



**UNIVERSIDAD CENTRAL DEL ECUADOR**

**FACULTAD DE INGENIERÍA, CIENCIAS FÍSICAS Y MATEMÁTICA**

**CARRERA DE INGENIERÍA INFORMÁTICA**

***ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DE UN DATA MART  
ACADÉMICO USANDO TECNOLOGÍA DE BI PARA LA FACULTAD DE  
INGENIERÍA, CIENCIAS FÍSICAS Y MATEMÁTICA.***

**Trabajo de Graduación Previo a la Obtención del Título de Ingeniero  
Informático**

**Autor:** Aimacaña Quilumba Doris Eulalia

**Tutor:** Ing. Javier Goyes

Quito – Ecuador  
2013

## **DEDICATORIA**

Dedico el presente trabajo de graduación a Dios por darme la fuerza que necesite en los momentos más difíciles de mi vida, a mis padres que con su sacrificio me apoyaron siempre en cada una de mis metas, a mis hermanos que a su modo me apoyaron y a todas las personas que ayudaron a culminar con este proyecto.

.

**Doris Aimacaña**

## **AGRADECIMIENTO**

Agradezco la valiosa colaboración del Ing. Javier Goyes como director de tesis por su apoyo y guía durante el desarrollo de la misma, a todos los catedráticos que desempeñan sus funciones en la Facultad de Ingeniería en Ciencias Físicas y Matemática, a la biblioteca y sus colaboradores que me apoyaron en el desarrollo académico, de igual forma a la Ing. Karina Serrano por su apoyo para desarrollar el proyecto.

**Doris Aimacaña**

## **AUTORIZACIÓN DE AUTORÍA INTELECTUAL**

Yo, Doris Eulalia Aimacaña Quilumba en calidad de autora del trabajo de investigación o tesis realizada sobre ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DE UN DATA MART ACADÉMICO USANDO TECNOLOGÍA DE BI PARA LA FACULTAD DE INGENIERÍA, CIENCIAS FÍSICAS Y MATEMÁTICA, por la presente autorizo a la UNIVERSIDAD CENTRAL DEL ECUADOR, hacer uso de todos los contenidos que me pertenecen o de parte de los que contiene esta obra, con fines estrictamente académicos o de investigación.

Los derechos que como autor me corresponden, con excepción de la presente autorización, seguirán vigentes a mi favor, de conformidad con lo establecidos en los artículos 5, 6, 8, 19 y demás pertinentes de la ley de Propiedad Intelectual y su Reglamento.

Quito, a 28 de Enero 2013



Doris Eulalia Aimacaña Quilumba

FIRMA

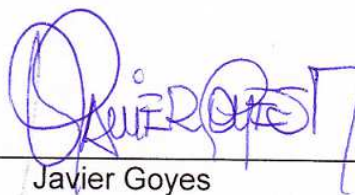
CC. 1719389171

## CERTIFICACION

En calidad de Tutor del proyecto de tesis: 'Análisis, Diseño e Implementación de un Data Mart Académico Usando Tecnología Bi para la Facultad de Ingeniería, Ciencias Físicas y Matemática, presentado y desarrollado por la señorita: Doris Eulalia Aimacaña Quilumba, previo a la obtención del Título de Ingeniero Informático considero que el proyecto reúne los requisitos necesarios.

En la ciudad de Quito a los 08 Días del mes de Enero del 2013

Atentamente,



Javier Goyes



UNIVERSIDAD CENTRAL DEL ECUADOR  
FACULTAD DE INGENIERÍA, CIENCIAS FÍSICAS Y MATEMÁTICA  
DIRECCIÓN CARRERA DE INGENIERÍA INFORMÁTICA

Oficio N° 38 -DC- IINF  
Quito, D.M., 17 de enero de 2013

Señores  
Ing. Rene CARRILLO  
Ing. Jorge GORDILLO  
Presente.

Señores Profesores:

A fin de dar cumplimiento a lo dispuesto en el "Reglamento para la Obtención de los Títulos Profesionales en la Facultad de Ingeniería, Ciencias Físicas y Matemática", aprobado por el H. Consejo Universitario, en sesión del 31 de octubre de 2011; agradeceré a usted, **calificar el Trabajo de Graduación** titulado: **"ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DE UN DATAMART ACADÉMICO USANDO TECNOLOGÍA DE BI PARA LA FACULTAD DE INGENIERIA , CIENCIAS FISICAS Y MATEMATICA** De la estudiante Doris Eulalia AIMACAÑA QUILUMBA, requisito previo a la obtención del título de **INGENIERA INFORMÁTICA**, en base al **Formulario del Resultado del Trabajo de Graduación**, que me permito remitirle.

Este formulario, deberá enviarse a la Secretaría General de la Facultad en un plazo no mayor a **ocho días**.

Atentamente,

Ing. Santiago Morales C. MSc.  
**DIRECTOR DE CARRERA  
INGENIERÍA INFORMÁTICA**



Recibi conforme

  
Ing. Rene Carrillo  
Ing. Jorge Gordillo

ELABORADO POR	CARGO	FIRMA	FECHA
Nancy Solís A.	Ayudante de Sec 2		17/01/2013



UNIVERSIDAD CENTRAL DEL ECUADOR  
FACULTAD DE INGENIERÍA, CIENCIAS FÍSICAS Y MATEMÁTICA  
DIRECCIÓN CARRERA DE INGENIERÍA INFORMÁTICA

## RESULTADO DEL TRABAJO DE GRADUACION

CARRERA DE: INGENIERÍA INFORMÁTICA

Quito, 18 Enero 2013

Señor (ita) DORIS EULALIA AIMACAÑA QUILUMBA

TEMA: "ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DE UN DATAMART ACADÉMICO USANDO  
TECNOLOGÍA DE BI PARA LA FACULTAD DE INGENIERIA, CIENCIAS FISICAS Y  
MATEMATICA"

### CALIFICACIÓN:

TRIBUNAL	PROFESOR (A)	NOTA SOBRE VEINTE		FIRMA
		NUMERO	LETRAS	
PROFESOR TITULAR	ING. RENE CARRILLO	20	veinte	[Firma]
PROFESOR TITULAR	ING. JORGE GORDILLO	18,0	Dieciocho	[Firma]
PROMEDIO		19	Diecinueve	[Firma]

Dra. Katheryne Carrión Valdivieso  
SECRETARIA ABOGADA, (E)

## Contenido

CAPITULO I.....	1
1. PRESENTACIÓN DEL PROBLEMA .....	1
1.1 Planteamiento del Problema.....	1
1.2 Formulación del Problema .....	3
1.3 Interrogantes de la Investigación .....	3
1.4 Objetivos de la Investigación.....	4
1.4.1 Objetivo General.....	4
1.4.2 Objetivos Específicos .....	4
1.5 Justificación .....	5
1.6 Alcance .....	5
1.7 Limitaciones.....	6
CAPITULO II.....	8
2 REVISIÓN BIBLIOGRÁFICA .....	8
2.1 Antecedentes .....	8
2.2 Fundamentación Teórica .....	8
2.2.1 Inteligencia de Negocios (Business Intelligence) .....	8
2.2.2 Pentaho .....	19
2.2.3 Pentaho PDI (Pentaho Data Integration) .....	22
2.2.4 Pentaho PSW (Pentaho Schema Workbench) .....	23
2.2.5 Pentaho PRD (Pentaho Report Designer).....	24
2.3 Identificación de Variables.....	25
2.4 Hipótesis .....	25
CAPITULO III.....	26
3 MARCO METODOLOGICO .....	26
3.1 Diseño de la Investigación .....	26
3.2 Diseño Experimental.....	26
3.3 Plan de Recolección de Datos. ....	27
3.4 Plan de Procesamiento de Datos .....	27
3.5 Metodología de Desarrollo .....	28
CAPITULO IV.....	29



4	DISEÑO DE LA SOLUCIÓN.....	29
4.1	Descripción de la Arquitectura.....	29
4.2	Análisis de las fuentes de datos .....	31
4.3	Calidad de Datos.....	32
4.4	Frecuencia de Carga .....	34
4.5	Modelado Multidimensional.....	34
4.5.1	Esquema Multidimensional.....	40
4.6	Diseño relacional de la base que soporta a los cubos .....	43
	CAPITULO V.....	50
5	IMPLEMENTACION DEL PROTOTIPO .....	50
5.1	Componentes de la arquitectura de la suite de Pentaho .....	50
5.2	Extracción, Transformación y Carga .....	51
5.2.1	Introducción .....	51
5.2.2	Procesos ETL.....	52
5.3	Cubos y Análisis OLAP .....	74
5.3.1	Estructura General .....	75
5.3.2	Publicación Cubos.....	85
5.4	Reportes e Indicadores .....	85
5.4.1	Estructura de los Reportes e Indicadores .....	86
5.5	Administración de Usuarios y Permisos.....	93
5.6	BI Server.....	93
	CAPITULO VI.....	95
6	CONCLUSIONES .....	95
6.1	Resultados.....	95
6.2	Conclusiones .....	95
6.3	Recomendaciones .....	96
	MATERIALES DE REFERENCIA .....	98
	BIBLIOGRAFÍA.....	98
	ANEXOS .....	100
	Anexo A.....	101
	Anexo B.....	102
	Anexo C.....	103

Anexo D. ....	107
Anexo E. ....	112

## Lista de Figuras

Figura 1. Beneficios BI .....	10
Figura 2. Gráfico de Gartner 2012 .....	11
Figura 3. Flujo de Información en BI .....	12
Figura 4. Proceso ETL .....	14
Figura 5. Data Mart dependiente .....	15
Figura 6. Data Mart Independiente .....	16
Figura 7. Esquema Estrella y Esquema Copo de Nieve .....	19
Figura 8. Arquitectura Pentaho .....	20
Figura 9. Report Designer .....	24
Figura 10. Arquitectura de la Solución .....	29
Figura 11. Vistas del Sistema Fing. ....	32
Figura 12. Dimensión Tiempo .....	35
Figura 13. Dimensión Periodo Lectivo .....	35
Figura 14. Dimensión Categoría Docente .....	36
Figura 15. Dimensión Dedicación Docente .....	36
Figura 16. Dimensión Facultad .....	37
Figura 17. Dimensión Especialidad .....	37
Figura 18. Dimensión Estado Estudiante .....	38
Figura 19. Dimensión Profesor .....	38
Figura 20. Dimensión Estudiante .....	39
Figura 21. Dimensión Estado Curso Estudiante .....	39
Figura 22. Dimensión Curso .....	40
Figura 23. Relación dimensional para el cubo de Docentes .....	40
Figura 24. Relación dimensional para el cubo de Estudiantes .....	41
Figura 25. Relación dimensional para el cubo de Docentes – Estudiantes .....	41
Figura 26. Relación dimensional para el cubo de Registro Académico .....	42
Figura 27. Relación dimensional para el cubo de la Cohorte .....	42
Figura 28. Relación dimensional para el cubo de Deserción .....	43
Figura 29. Modelo Multidimensional en base a tablas relacionales para el cubo de DOCENTES .....	44
Figura 30. Modelo Multidimensional en base a tablas relacionales para el cubo de ESTUDIANTES .....	45
Figura 31. Modelo Multidimensional en base a tablas relacionales para el cubo de DOCENTE - ESTUDIANTE .....	46
Figura 32. Modelo Multidimensional en base a tablas relacionales para el cubo de REGISTRO ACADEMICO .....	47
Figura 33. Modelo Multidimensional en base a tablas relacionales para el cubo de la COHORTE .....	48

Figura 34. Modelo Multidimensional en base a tablas relacionales para el cubo de la DESERCIÓN .....	49
Figura 35. Proceso ETL.....	51
Figura 36. Transformación STG_CARGA_HORARIA.....	52
Figura 37. Transformación STG_ESTUDIANTE_INDICADORES .....	52
Figura 38. Transformación STG_EST_DOC_MAT .....	53
Figura 39. Trabajo Job_STG.....	53
Figura 40. Transformación STG_DIM_CATEGORIA_DOCENTE .....	54
Figura 41. Transformación STG_DIM_DEDICACIÓN_DOCENTE.....	54
Figura 42. Transformación STG_DIM_ESTADO_ESTU_CURSO .....	54
Figura 43. Transformación STG_DIM_ESTADO_ESTUDIANTE .....	55
Figura 44. Transformación STG_DIM_ESTUDIANTE .....	55
Figura 45. Transformación STG_DIM_FACULTAD.....	55
Figura 46. Transformación STG_DIM_PERIODO_LLECTIVO .....	56
Figura 47. Transformación STG_DIM_PROFESOR.....	56
Figura 48. Transformación STG_DIM_CURSO .....	56
Figura 49. Transformación STG_DIM_ESPECIALIDAD .....	56
Figura 50. Trabajo Job_STG_Dimensiones .....	57
Figura 51. Transformación Dim_Fecha.....	58
Figura 52. Transformación DIM_CATEGORIA_DOCENTE.....	58
Figura 53. Transformación DIM_DEDICACION_DOCENTE.....	59
Figura 53. Transformación DIM_ESTADO_ESTU_CURSO .....	59
Figura 54. Transformación DIM_ESTADO_ESTUDIANTE.....	59
Figura 55. Transformación DIM_ESTUDIANTE.....	60
Figura 56. Transformación DIM_FACULTAD.....	60
Figura 57. Transformación DIM_PERIODO_LLECTIVO.....	60
Figura 58. Transformación DIM_PROFESOR .....	60
Figura 59. Transformación DIM_CURSO.....	61
Figura 60. Transformación DIM_ESPECIALIDAD .....	61
Figura 61. Trabajo Job_Dimensiones .....	61
Figura 62. Transformación TMP_ESTUDIANTE .....	62
Figura 63. Transformación FC_ESTUDIANTE_DETALLE .....	62
Figura 64. Transformación FC_ESTUDIANTE .....	63
Figura 65. Trabajo Job_Fc_Estudiente .....	63
Figura 66. Transformación TMP_DOCENTE .....	63
Figura 67. Transformación FC_DOCENTE_DETALLE .....	64
Figura 68. Transformación FC_DOCENTE .....	64
Figura 69. Trabajo Job_Fc_Docente.....	65
Figura 70. Transformación TMP_DOCENTE-ESTUDIANTE .....	65
Figura 71. Transformación FC_ESTUDIANTE-DOCENTE_DETALLE .....	65
Figura 72. Transformación FC_ESTUDIANTE_DOCENTE.....	66
Figura 73. Trabajo Job_Fc_Est_Doc .....	66

Figura 74. Transformación TMP_REGISTRO_ACADEMICO .....	66
Figura 75. Transformación FC_REGISTRO_ACADEMICO_DETALLE .....	67
Figura 76. Transformación FC_REGISTRO_ACADEMICO .....	67
Figura 77. Trabajo Job_Fc_Registro_Academico .....	68
Figura 78. Transformación TMP_COHORTE.....	68
Figura 79. Transformación FC_COHORTE_DETALLE .....	68
Figura 80. Transformación FC_COHORTE .....	69
Figura 81. Trabajo Job_Fc_Cohorte .....	69
Figura 82. Transformación TMP_COHORTE_PERIODO_1.....	69
Figura 83. Transformación TMP_COHORTE_PERIODO_2.....	70
Figura 84. Transformación TMP_COHORTE_PERIODO_3.....	70
Figura 85. Trabajo Job_Tmp_Periodo .....	70
Figura 86. Transformación TMP_DESERCION .....	71
Figura 87. Transformación FC_DESERCION_DETALLE.....	71
Figura 88. Transformación FC_DESERCION.....	71
Figura 89. Trabajo Job_Fc_Desercion.....	72
Figura 90. Representación Cubo Docentes .....	76
Figura 91. Representación Cubo Docentes Detallado .....	77
Figura 92. Representación Cubo Estudiantes.....	77
Figura 93. Representación Cubo Estudiante Detallado .....	78
Figura 94. Representación Cubo Docente-Estudiante .....	79
Figura 95. Representación Cubo Docente-Estudiante Detallado .....	80
Figura 96. Representación Cubo Registro Académico .....	80
Figura 97. Representación Cubo Registro Académico Detallado .....	81
Figura 98. Representación Cubo Cohorte .....	82
Figura 99. Representación Cubo Cohorte Detallado .....	83
Figura 100. Representación Cubo Deserción.....	83
Figura 101. Representación Cubo Deserción Detallado .....	84
Figura 102. Reporte Horario Docente.....	87
Figura 103. Reporte Estudiante por Docente .....	88
Figura 104. Indicadores Carga Horaria Cuerpo Docente.....	89
Figura 105. Indicadores Docente.....	91
Figura 106. Indicadores Docente.....	92
Figura 107. Pestaña de gestión de usuarios en la consola de administración.....	93
Figura 108. Sectores de trabajo de la Consola de Usuario Pentaho .....	94

## Lista de Tablas

Tabla 1. Problemas que soluciona el BI .....	10
Tabla 2. Cuadro comparativo de herramientas de extracción .....	22
Tabla 3. Mapeo DIM_FECHA.....	107
Tabla 4. Mapeo DIM_PERIODO_LECTIVO.....	108
Tabla 5. Mapeo DIM_CATEGORIA_DOCENTE .....	108
Tabla 6. Mapeo DIM_DEDICACION_DOCENTE.....	108
Tabla 7. Mapeo DIM_ESTADO_ESTUDIANTE .....	108
Tabla 8. Mapeo DIM_FACULTAD.....	109
Tabla 9. Mapeo DIM_ESPECIALIDAD.....	109
Tabla 10. Mapeo DIM_ESTADO_ESTU_CURSO .....	110
Tabla 11. Mapeo DIM_CURSO.....	110
Tabla 12. Mapeo DIM_PROFESOR.....	110
Tabla 13. Mapeo DIM_ESTUDIANTE.....	111
Tabla 14. Mapeo FC_DOCENTE_DET .....	113
Tabla 15. Mapeo FC_DOCENTE .....	114
Tabla 16. Mapeo FC_ESTUDIANTE_DET .....	115
Tabla 17. Mapeo FC_ESTUDIANTE .....	116
Tabla 18. Mapeo FC_REG_ACADEMICO_DET .....	117
Tabla 19. Mapeo FC_REGISTRO_ACADEMICO .....	118
Tabla 20. Mapeo FC_COHORTE_DET .....	119
Tabla 21. Mapeo FC_COHORTE .....	120
Tabla 22. Mapeo FC_DESERCION_DET .....	121
Tabla 23. Mapeo FC_DESERCION_DET .....	122

## **RESUMEN**

### **ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DE UN DATA MART ACADÉMICO USANDO TECNOLOGÍA DE BI PARA LA FACULTAD DE INGENIERÍA, CIENCIAS FÍSICAS Y MATEMÁTICA**

La Facultad de Ingeniería requiere unos indicadores de gestión para la toma de decisiones estratégicas tanto de estudiantes como profesores, esto es muy importante para saber dónde y cómo solventar alguna falla, o mejorar las decisiones tomadas.

Para darle una solución al problema se propone realizar un Data Mart que nos permite almacenar sólo la información requerida e ir eliminando aquellos datos que obstaculizan la labor del análisis y entregando la información que se requiera en la forma más apropiada, facilitando así el proceso de gestión de la misma.

Los Data Mart surgen con la promesa del manejo y control de la información, aseguran una vista única de los datos, que pueden provenir de diversas fuentes.

Existen diferentes herramientas BI tanto software libre como propietario, para esta solución por un requisito de la universidad se usara software libre y después de un análisis en cuanto a herramientas de este tipo se usó Pentaho.

#### **DESCRIPTORES:**

INDICADORES / TOMA DE DECISIONES / ESTRATEGIAS / DATA MART /BI  
/SOFTWARE LIBRE /PENTAHO

## **ABSTRACT**

### **ANALYSIS, DESIGN AND IMPLEMENTATION OF ACADEMIC DATA MART USING TECHNOLOGY FOR BI FOR "FACULTAD DE INGENIERÍA CIENCIAS FÍSICAS Y MATEMÁTICA"**

The "Facultad de Ingeniería Ciencias Físicas y Matemática" requires management indicators for strategic decision making of both students and teachers, this is very important to know where and how to remedy any failure, or improve decisions. To give a solution proposed to develop a Data Mart allows us to store only the information required and to eliminate those that hamper data analysis and submitting the required information in the most appropriate way, facilitating the management process thereof.

The Data Mart come with the promise of the management and control of information, ensure a single view of data, which may come from various sources. There are different BI tools both free and proprietary for this solution for a requirement of the university will use free software after analysis regarding such tools used Pentaho.

#### **DESCRIPTORS:**

INDICATORS / DECISION MAKING / STRATEGIES / DATA MART / BI / FREE  
SOFTWARE / PENTAHO



## INTRODUCCIÓN

El sistema académico de la Facultad de Ingeniería maneja miles de registros de datos sobre estudiantes, docentes y empleados, estos datos podrían ser aprovechados para ayudar en la toma de decisiones, mejorando el manejo de esta información en las distintas áreas donde uno de los principales objetivos es integrar y compartir información asignándoles varios atributos que sirvan como base para el análisis.

Con la informatización de la sociedad y dentro de esta las instituciones educativas han crecido a pasos agigantados, así como la capacidad de generación y almacenamiento de la información que no puede ser analizada por los métodos tradicionales existentes, mientras mayor es la capacidad para almacenar más y más datos, mayor es la incapacidad para extraer información realmente útil de éstos, motivo por el cual mucha información importante, queda sepultada y disgregada.

Los sistemas existentes no estaban preparados para un análisis a fondo. Mucho se ha hablado de la “Era de la Información” y sus ventajas; con las nuevas posibilidades se acortan las distancias y crecen los beneficios para quienes tienen acceso a la gran cantidad de datos. Sin embargo, lo que constituye un valioso recurso para todos, se ha tornado en un gran problema, la solución es manejar de forma óptima grandes volúmenes de información.

La competencia en el nuevo ambiente educativo donde todo está marcado por la información, el conocimiento del mercado y la toma de decisiones es muy importante para saber dónde y cómo se organiza toda la información. En este contexto surgen términos que se convierten en conceptos y filosofías de trabajo, como Data Mart.

Un Data Mart aporta a la institución, eliminando aquellos datos que obstaculizan la labor del análisis de información y entregan la información que se requiera en la forma más apropiada, facilitando así el proceso de gestión de la misma.

Los Data Mart surgen con la promesa del manejo y control de la información, aseguran una vista única de los datos, que pueden provenir de diversas fuentes.

El objetivo de utilizar Inteligencia de Negocios en nuestra facultad para la construcción del Data Mart, es colocar los datos al alcance de los responsables de la toma de decisiones, utilizando herramientas que extraigan los datos de la mejor manera para posteriormente almacenarlos en un repositorio optimizando la entrega de información de forma rápida y resumida que haga posible un análisis detallado y completo de los indicadores que muestren el desempeño y comportamiento de las actividades desarrolladas en la facultad, así como la información referente al ámbito académico de los estudiantes y docentes.

## CAPITULO I

### 1. PRESENTACIÓN DEL PROBLEMA

#### 1.1 Planteamiento del Problema

Actualmente la facultad no cuenta con un sistema que permita la toma de decisiones de manera proactiva, los procesos que se manejan en la facultad, cada día son más complejos y su gestión se convierte más difícil por diversas razones. En tanto las tecnologías de la información en la actualidad permiten poder mejorar este desempeño organizacional desde la parte operativa hasta la etapa gerencial para la toma de decisiones de las cuales depende toda la organización en este caso nuestra facultad.

Se cuenta con la información académica de estudiantes y docentes almacenada en una base de datos, pero no se ha pensado en que esta información se la puede convertir en conocimiento, de modo que pueda servirle a las autoridades de la facultad para crear planes de estrategia y toma de decisiones basadas en los resultados obtenidos al consolidar la información en un Data Mart y poder presentar esto en Cubos de Información, Indicadores y Reportes fácil de interpretar y manejar.

En relación con los problemas identificados se presenta el planteamiento de la solución destinado a mejorar la toma de decisiones educativas estratégicas para la Facultad:

*Análisis, Diseño e Implementación de un Data Mart Académico usando Tecnología de BI para la Facultad de Ingeniería, Ciencias Físicas y Matemática.*

La solución empezaría por la consolidación de la información en un Data Mart, tomando en cuenta los requerimientos solicitados por las personas que manejan esta información y requieren poder analizarla con mayor facilidad y tener acceso a información adecuada, integrada y actualizada.

Para el manejo de esta información se utilizará herramientas Open Source como son: Pentaho, que permitirá realizar la creación del Data Mart, mediante el proceso de Extracción, Transformación y Carga de la información (ETL) del cual se podrá obtener Cubos de información para el análisis de las distintas necesidades de las autoridades y la creación de consultas mostradas mediante reportes e indicadores para poder observar el desempeño de la Facultad y realizar la toma de decisiones mediante el análisis de los indicadores de gestión y su comportamiento.

### **Macro**

En las grandes empresas se maneja una gran cantidad de información y el reto es saber administrarla. Por ejemplo, el director general, necesita información para tomar una decisión, por lo que solicita al asistente un informe y este a su vez solicita al director de área el informe y este al área de sistemas, realizar este proceso toma mucho tiempo ya que la mayor parte del análisis se lo realiza de forma manual, usar una herramienta que nos permita minimizar el flujo de trabajo para obtener una fuente dinámica y obtener así un reporte gerencial, permite optimizar tiempos obteniendo como resultado información mejor administrada y fácil de usar.

### **Meso**

En las Universidades se maneja gran cantidad de datos de las distintas áreas que existen de los cuales se puede obtener buenos Data Mart, usándolos para generar reportes e indicadores que permitan un manejo y toma de decisiones que ayuden en áreas donde se busca la reducción de costos, tiempos y esfuerzos. Entregar soluciones a los consumidores de información en forma simple y entendible sin necesidad de realizar análisis complicados que toman mucho tiempo, implicando un desarrollo flexible y un diseño ajustado a cada necesidad, es decir justo a la medida de cada uno de los usuarios.

## **Micro**

Actualmente los directores de carrera de la Facultad de Ingeniería requieren ciertos indicadores sobre estudiantes y docentes para la evaluación interna y externa de las carreras, el manejo de la gran cantidad de datos es la problemática ya que en la actualidad no se cuenta con las herramientas necesarias para que esto se lo haga de forma automática, actualmente se despliegan en documentos de Excel creados de forma manual, la Inteligencia de Negocio es una gran opción para la automatización de estos procesos logrando optimizar tiempo y manejo de la información mediante la creación del Data Mart, Reportes e Indicadores.

## **Prognosis**

La Implementación de un Data Mart Académico usando Tecnología de BI para la Facultad de Ingeniería, Ciencias Físicas y Matemática mejoraría el flujo del manejo de los datos, convirtiéndolos en información para una buena toma de decisiones estratégicas.

### **1.2 Formulación del Problema**

¿Cuál es la ventaja de realizar el Proyecto de Análisis, Diseño e Implementación de un Data Mart Académico usando Tecnología de BI para la Facultad de Ingeniería, Ciencias Físicas y Matemática?

### **1.3 Interrogantes de la Investigación**

- ¿Es factible la implementación de la tecnología BI para la Facultad de Ingeniería?
- ¿Se puede aplicar la plataforma tecnológica para la creación de la fuente de los indicadores de gestión que utilizaran los Directores de Carrera?
- ¿Se podrán integrar todas las fuentes de información, para la extracción y carga de todos los datos?

- ¿Es posible obtener toda la información requerida para la construcción de los Data Mart que serán fuente de los indicadores de gestión de los Directores de Carrera?
- ¿La herramienta utilizada para el proceso de extracción y carga de datos podrá sustentar los procesos necesarios para la construcción del Data Mart Académico para la facultad?
- ¿Los cubos de información cubren todos los requerimientos en cuanto a tiempo de ejecución?
- ¿Los indicadores serán una buena fuente de información para la toma de decisiones sobre la autoevaluación académica?

## **1.4 Objetivos de la Investigación**

### **1.4.1 Objetivo General**

Realizar el análisis, diseño e implementación de Data Mart Académicos usando Tecnología de BI para la Facultad de Ingeniería, Ciencias Físicas y Matemática, poniendo a disposición de las autoridades, la Información consolidada de modo que permita agilizar el proceso de análisis de datos, formulación de estrategias de prevención y planificación de actividades de una forma más rápida y eficaz.

### **1.4.2 Objetivos Específicos**

Identificar y evaluar los componentes de los Procesos Técnicos y Especializados, como herramienta para la implementación de Data Mart Académicos que se van a utilizar para cumplir con las necesidades de la Facultad.

- Fijar metas estratégicas y evaluar la gestión estratégica para la creación de los Data Mart.
- Realizar de manera eficiente el proceso de ETL (Extracción, Transformación y Carga de Datos)

- Lograr que el proceso de ETL, cubos, reportes e indicadores sea fácil de generar, mantener y desplegar.
- Analizar las diferentes fuentes de información para determinar la creación de la aplicación y los beneficios que proporciona al presentar la misma en Data Mart.
- Definir una estructura dimensional que permita una salida del análisis de datos obtenidos del Data Mart.
- Crear reportes e indicadores para la toma de decisiones de la Facultad.

### **1.5 Justificación**

Basado en los antecedentes anteriormente expuestos, la solución al problema sirve como apoyo a la Facultad de Ingeniería para cubrir sus necesidades de mejoramiento en el manejo de la información de modo que pueda ser más eficiente la toma de decisiones organizacionales de nivel operativo y estratégico.

Además se plantea la solución por parte del autor de la propuesta de tesis, para poder poner en práctica los conocimientos adquiridos en la Carrera de Ingeniería Informática en un ambiente real en beneficio de la sociedad.

La importancia del presente trabajo de investigación radica en agilizar y automatizar todo el proceso de gestión de la información con el objetivo de brindar un servicio de calidad con una herramienta que cumpla las expectativas de la Facultad, con ello eliminar las deficiencias en tiempo en el entorno administrativo y técnico con un manejo rápido y eficaz de la información, para ir acorde con el crecimiento de la misma.

### **1.6 Alcance**

La creación de Data Marts Académicos para la Facultad de Ingeniería, contendrá lo siguiente:

- Se creará los Data Mart necesarios, que contengan la información académica de la Facultad, para la creación de los indicadores de gestión, para la

evaluación externa e interna de las carreras que solicitaron los directores, obteniendo las mejores prácticas para la administración de esta información.

- Se crearán los Data Mart para los indicadores que tenga como fuente el sistema académico de la facultad, los demás indicadores aún no cuentan con una fuente de consulta, es decir aún no existe una fuente para esos datos.
- Implementar sobre la herramienta libre de Inteligencia de Negocios la solución para definir y administrar la información relacionada con los estudiantes y docentes de la Facultad.
- Como resultado se obtendrán los modelos dimensionales que permitirán la creación de Cubos OLAP, que serán la fuente de los Reportes e Indicadores que muestren la información para la toma de decisiones estratégicas de la Facultad.
- Se entregará toda la documentación referente a este proyecto de tesis para que la facultad pueda continuar cuando crea conveniente con el proceso de implementación, ya que se contó solamente con una fuente de datos limitada lo que no permitió realizar la implementación.

### **1.7 Limitaciones**

Es pertinente dar al problema una formulación lógica y adecuada, así como también precisar sus límites y alcance, para ello es necesario tener en cuenta los siguientes factores:

- La herramienta que se define para usar es una herramienta libre por lo que no se dispone de un soporte e información completo.
- La plataforma sobre la cual se va implementar la herramienta de Inteligencia de Negocios es en base a la plataforma actual con la que trabaja la facultad.
- La estructura de las distintas fuentes de información no es la misma, motivo por el cual hay que homologar dichas fuentes en caso de existir la homologación.



- El desarrollo de la Aplicación se realizó fuera del horario laboral, lo que limitó el trato directo y constante con el usuario.
- La disponibilidad de obtener los datos de las fuentes requeridas para la construcción de los indicadores.

## **CAPITULO II**

### **2 REVISIÓN BIBLIOGRÁFICA**

#### **2.1 Antecedentes**

En la actualidad existen varios proyectos realizados con tecnología BI (Inteligencia de Negocios) que sirven como apoyo para la toma de decisiones gerenciales de instituciones públicas y privadas, en nuestro País existen empresas que se encargan de realizar proyectos basados en el BI, uno de los cuales sirvió como ejemplo para el desarrollo de esta tesis es la “Creación del proceso de ETL para la elaboración de un Dashboard Académico para la Universidad de las Américas”; este proyecto se lo realizó en una herramienta BI pagada Cognos, se analizó como se realizó el levantamiento de información y procesos para la creación del ETL de este proyecto, como sus cubos, reportes e indicadores.

Actualmente en la facultad no existe información sobre la realización de algún aplicativo informático que cumpla con los requerimientos propuestos en el presente plan de tesis para el manejo de la información, para la creación de la fuente de los indicadores de gestión, basados en la tecnología de la Inteligencia de Negocios; motivo por el cual no existe una base bibliográfica en la facultad que sirva como base para el desarrollo de esta tesis.

#### **2.2 Fundamentación Teórica**

La Facultad de Ingeniería en Ciencias Físicas y Matemática requiere el uso de Data Mart Académicos que facilite la toma de decisiones estratégicas, a continuación se presenta un marco teórico referencial que presenta breves introducciones de conceptos que serán de utilidad para comprender los objetivos del presente proyecto de tesis.

##### **2.2.1 Inteligencia de Negocios (Business Intelligence)**

La Inteligencia de Negocios es la habilidad de una empresa para estudiar sus comportamientos y acciones históricas con el objetivo de entender donde la

organización ha estado, su situación actual, y predecir que sucederá en el futuro, es el proceso de recolectar, consolidar y analizar múltiples fuentes de datos para la toma de decisiones estratégicas.

En una sola palabra: "Análisis".

### **Beneficios de la implantación del BI**

Los primeros sistemas de información a la dirección aportaban básicamente información económica y financiera, con lo que quedaban muchas lagunas por cubrir respecto a la organización de la empresa. Los nuevos sistemas de BI abarcan todas las áreas funcionales de una organización (recursos humanos, marketing, logística,...) y no sólo los económicos.

Los beneficios que pueden aportar la implantación de sistemas BI son: incremento de la eficiencia en la toma de decisiones, mejora de comunicación entre las diferentes áreas de la organización, mejora de rentabilidad, atracción de nuevos clientes, reducción del fraude, identificación de riesgos, presentación de soluciones basadas en el conocimiento del negocio, etc.

La mayor parte de los beneficios de la implantación de un sistema de BI son intangibles, ya que derivan en la mejora de la gestión de la compañía. Esto dificulta la tarea de calcular su ROI (Return On Investment) para obtener beneficios cuantificables.

Gracias a la implantación de un sistema de BI, se evidencia una reducción de costes al aumentar el rendimiento de la infraestructura TIC de la organización y un incremento de la productividad de los empleados debido a la disponibilidad de información y la calidad de ésta. Pero como se ha comentado anteriormente este tipo de parámetros son difícilmente cuantificables desde una perspectiva económica-financiera.

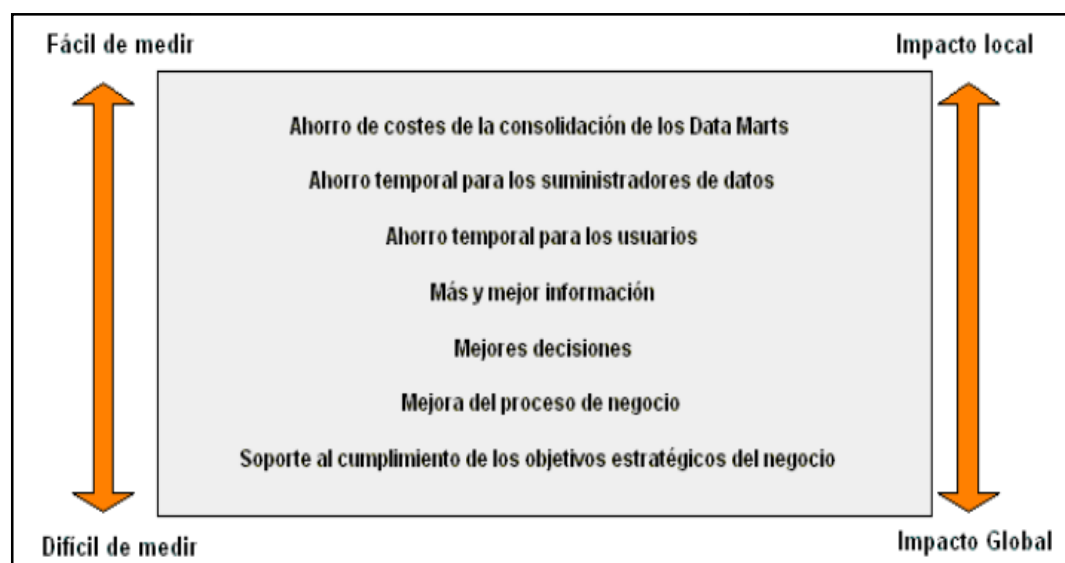


Figura 1. Beneficios BI

### Problemas que la Inteligencia de Negocios soluciona:

Banca & Seguros	Retail & Productos Empaquetados	Industrial	Empresas Publicas	Telcos
<ul style="list-style-type: none"> <li>• Porque nos dejan los clientes</li> <li>• Que clientes son rentables</li> <li>• Que ofertas puedo hacer para retener mis clientes</li> <li>• Que reclamos son fraudulentos</li> <li>• Como podemos encontrar nuevos clientes</li> </ul>	<ul style="list-style-type: none"> <li>• Excesivo o muy poco stock de inventario</li> <li>• Pobre abastecimiento de proveedores</li> <li>• Falta de estrategia de precios</li> <li>• Pobres iniciativas de fidelidad</li> <li>• Inhabilidad de seguir y medir los efectos de las campañas</li> </ul>	<ul style="list-style-type: none"> <li>• Cual fue el inventario pronosticado vs. el real</li> <li>• Que proveedores son efectivos</li> <li>• Cuanto es el excedente de inventario y porque</li> <li>• Como podemos optimizar las compras de inventario</li> </ul>	<ul style="list-style-type: none"> <li>• No hay presupuestos ni pronósticos consolidados</li> <li>• Pobre respuesta a las regulaciones</li> <li>• Evasión de impuestos</li> </ul>	<ul style="list-style-type: none"> <li>• Como podemos controlar y mantener a nuestros clientes</li> <li>• Como podemos reducir perdidas</li> <li>• Que ofertas podemos ofrecer a nuestros clientes para retenerlos a largo plazo de una manera rentable</li> </ul>

Tabla 1. Problemas que soluciona el BI

### Cuadrante mágico de la Inteligencia de Negocios

El cuadrante mágico es una herramienta analítica creada y promovida por la empresa Gartner y la cual muestra una representación gráfica del mercado compartido en un determinado periodo de tiempo. Los Cuadrantes Mágicos de

Gartner proporcionan a las empresas un medio para identificar y diferenciar a los proveedores de servicios del sector de las tecnologías de la información.

Según define Gartner, los líderes en los cuadrantes mágicos son aquellos fabricantes de software que operan bien hoy día, tienen una visión clara de la dirección del mercado y desarrollan activamente las competencias necesarias para mantener su posición de líderes en el mercado.

A continuación se muestra el cuadrante de Gartner del presente año.

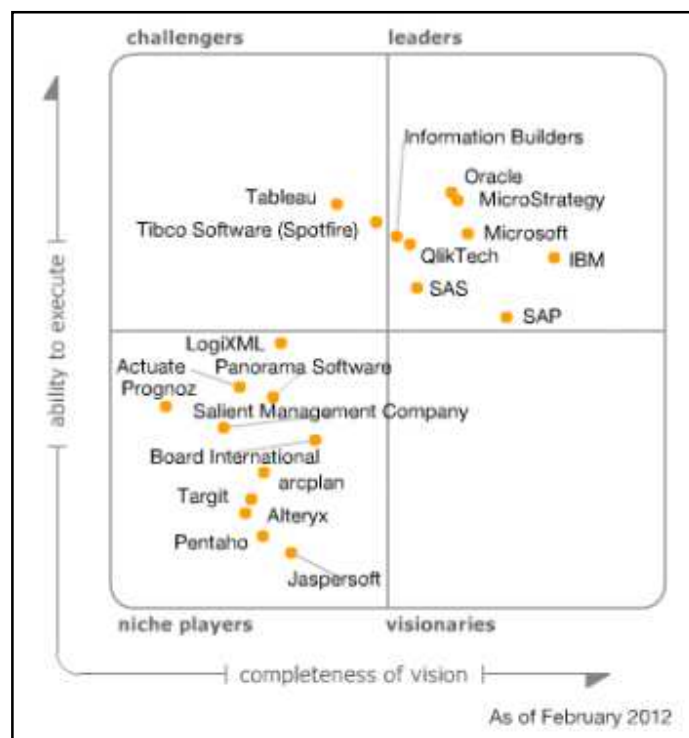


Figura 2. Gráfico de Gartner 2012

## Componentes y Arquitectura del BI

### ■ Sistemas Operacionales

- Son los sistemas operacionales de registros que capturan las transacciones del negocio.
- Pueden ser muchos sistemas con diferentes bases de datos y/o archivos planos.

- ETL

- Los datos son enviados a través de un proceso de extracción, transformación y carga (ETL) a una área para ser limpiado y cambiado antes de cargarlos en el data warehouse. Esta área es conocida como Staging Area. Su principal labor es evitar que el proceso de limpiado se haga en el momento que se carga los datos de los sistemas operacionales.

- Data Warehouse

- Es el repositorio de la organización.
- La información se califica en base a los intereses de una empresa.
- La información contenida en un Data Warehouse se encuentra integrada, lo que permite que los datos se puedan visualizar de diversas maneras.

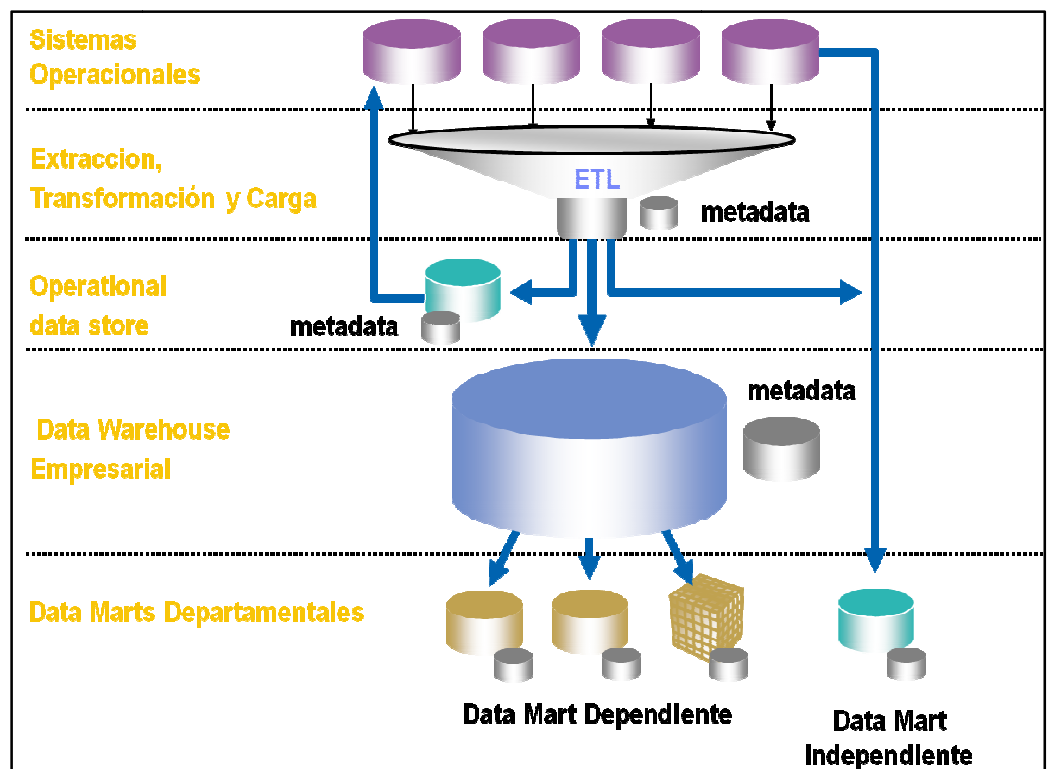


Figura 3. Flujo de Información en BI

- Data Marts
  - Hay dos tipos de data marts:
    - **Dependientes:** que obtiene sus datos del data warehouse.
    - **Independientes:** que obtiene sus datos de fuentes separadas.
- Fuente de Datos Operacional (Operational Data Source ODS)
  - Es usado para la toma de decisiones tácticas.
  - Tiene una historia mínima y muestra el estado de la entidad tan cerca del tiempo real como le sea posible.
  - El dato es volátil o actualizable.
- Cubos
  - Son estructuras multidimensionales.
  - Contiene datos pre-calculados.
- Metadata
  - Es la descripción de qué es cada campo, de dónde viene, y cómo es usado.
    - **Metadata técnico:** describe la estructura física y el proceso que mueve y transforma datos en el ambiente.
    - **Metadata de negocio:** describe la estructura de datos, reglas del negocio.
- Reportes
  - Recopilan los datos de las tablas o consultas para permitir su impresión o análisis, facilitando la individualización de los datos más importantes y su representación.

### 2.2.1.1 Proceso ETL

ETL este término viene de inglés de las siglas Extract-Transform-Load que significan Extraer, Transformar y Cargar y se refiere a los datos en una empresa. ETL es el proceso que organiza el flujo de los datos entre diferentes sistemas en una organización y aporta los métodos y herramientas necesarias para mover datos desde múltiples fuentes a un almacén de datos, reformatearlos, limpiarlos y cargarlos en otra base de datos, data mart ó bodega de datos. ETL forma parte de la Inteligencia Empresarial (Business Intelligence), también llamado “Gestión de los Datos” (Data Management).

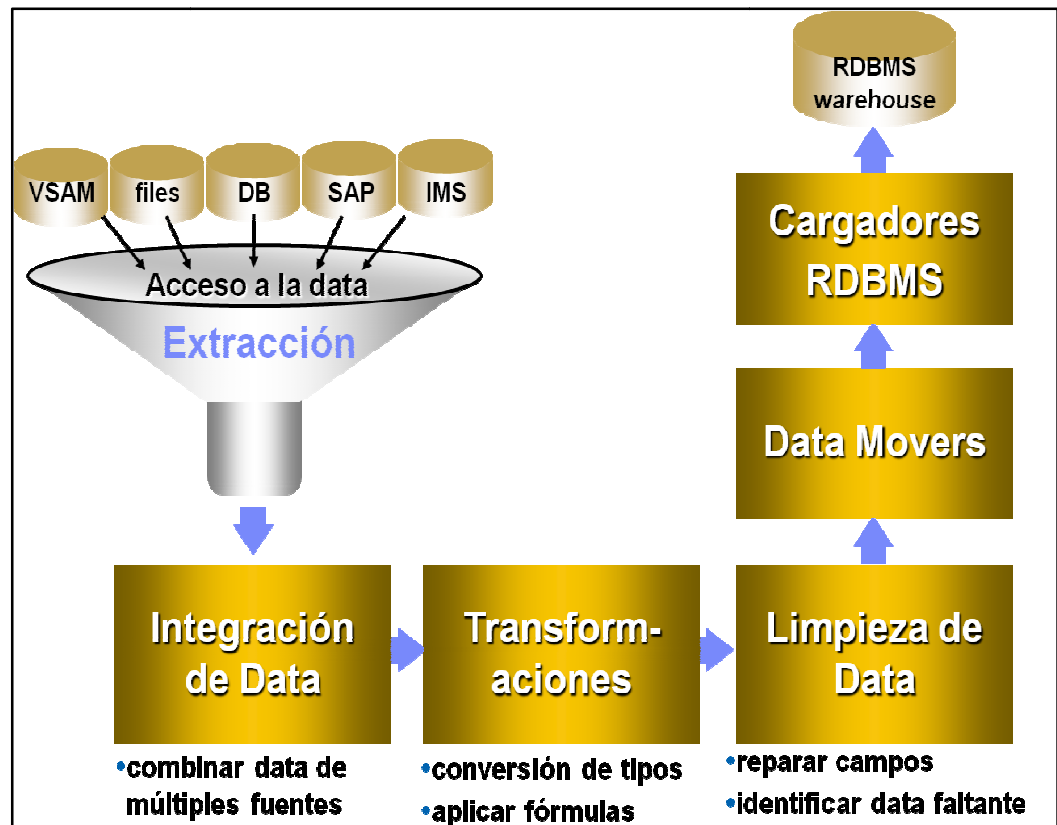


Figura 4. Proceso ETL

#### Data Warehouse:

Un Data Warehouse, o almacén de datos, es la colección de datos en donde se encuentra la información de forma integrada de una determinada institución, con el propósito de facilitar las tomas de decisiones para esta, ya que se encuentran



todos los elementos de los datos desde diversas fuentes de integración en un ambiente de aplicación, simplificando el tiempo y proceso de búsqueda.

Todas las herramientas para la toma de decisiones que se basan en un Data Warehouse, hacen más práctica y fácil la explotación de los datos, esto no se logra usando los datos de las aplicaciones operacionales (operaciones cotidianas), en donde la información se obtiene mediante procesos independientes y complejos

### **Data Mart:**

Un Data Mart es una versión especial de almacén de datos. Son subconjuntos de datos con el propósito de ayudar a que un área específica dentro del negocio pueda tomar mejores decisiones. Los datos existentes pueden ser utilizados por múltiples grupos de usuarios dependiendo de sus necesidades.

### **Tipos de Data Mart**

Existen dos tipos de Data Mart, los dependientes e independientes:

Dependientes: Son los que se construyen a partir de un Data Warehouse central, es decir reciben sus datos de un repositorio empresarial central.

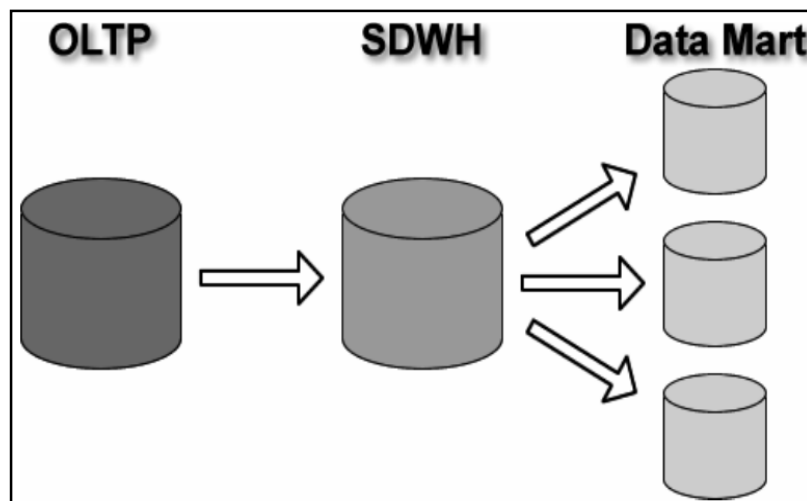


Figura 5. Data Mart dependiente

Independientes: Son aquellos Data Mart que no dependen de un Data Warehouse central, ya que pueden recibir los datos directamente del ambiente operacional, ya sea mediante procesos internos de las fuentes de datos o de almacenes de datos operacionales (ODS).

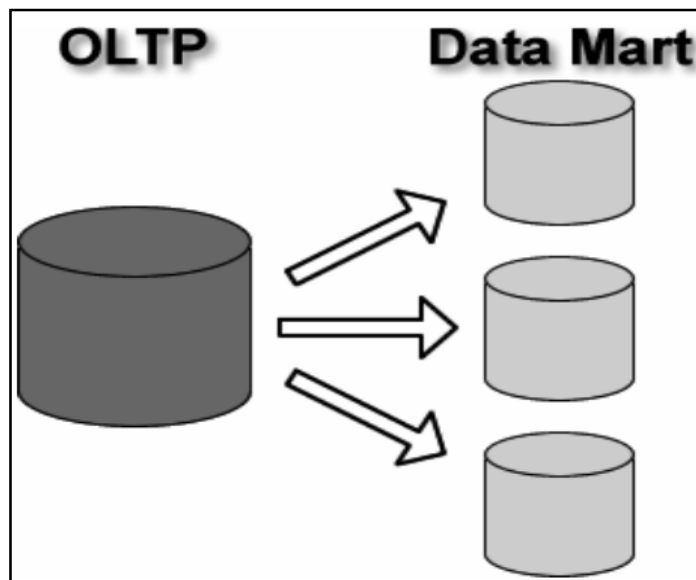


Figura 6. Data Mart Independiente

### **Carga de datos en un Data Mart**

Para la carga de datos hacia el Data Mart se pueden utilizar técnicas de carga para las herramientas OLAP, pero se debe tener en cuenta la capacidad para soportar la extracción de gran volumen de datos desde las fuentes, para no sobrecargar las mismas. Además los tiempos de carga y la calidad de los datos a ser cargado en el Data Mart.

Existen algunas herramientas para realizar la carga de datos libres y propietarios.

Se debe tener cuidado que los datos sean coherentes es decir que los datos que son transportados sean los que se cargan.

Esta fase comprende: preparación, integración, alto nivel de agregación y personalización de datos.

### **Comparación de Data Marts v/s Data Warehouse**

La implementación es muy similar a la de un Data Warehouse, con funcionalidades similares. Necesita de los mismos recursos como si corriera sobre un Data Warehouse, por lo tanto necesita cantidad de recursos similares.

- No se tienen todos los datos de la empresa, contiene los datos de un determinado sector de la empresa, por lo que las consultas tardan lo mismo que al realizarlas a un Data Warehouse.
- Actualizar el data mart desde el data warehouse cuesta menos, ya que los formatos de los datos son o suelen ser idénticos.

#### **2.2.1.2 Tecnología OLAP**

Provee una alternativa a la bases de datos relacionales, ofreciendo velocidad, flexibilidad de ver la información, análisis y navegación.

##### **Características:**

- Tiene un esquema optimizado para consultas de usuarios más rápido.
- Tiene un robusto motor de cálculo para análisis numéricos.
- OLAP provee una vista de datos multidimensional.
  - Drill
  - Pivot

##### **Cubos OLAP (OnLine Analytical Processing o Procesamiento Analítico en Líne)**

Los cubos de información, o cubos OLAP, son herramientas que permiten una visión multidimensional de los datos, y debido a su número indefinido de dimensiones, también se les llaman hipercubos.

Está compuesto por dimensiones y variables. Las dimensiones son atributos de las variables, información complementaria que se necesita para presentar los datos a

los usuarios, como nombres, zonas, descripciones, etc. Las variables, o indicadores de gestión, son los datos analizados. Representan un aspecto medible de los objetos o eventos a analizar. Por lo general se les representan con valores detallados para cada instancia del objeto medido.

Para obtener cierta información, el usuario hace una consulta seleccionando los atributos que desea ver, y las restricciones, como por ejemplo, determinado período de tiempo, determinado producto, etc. Sin embargo, lo que hace de esta herramienta que sea útil son sus operadores de refinamiento, drill, el cual permite agregar un nuevo criterio de agrupación, roll, que permite eliminar un criterio de agrupación, slice & dice, que permite seleccionar datos para presentarlos en un informe, y pivot, el cual permite reorientar las dimensiones del cubo.<sup>1</sup>

### **Esquema Estrella y Copo de Nieve**

Al modelar las dimensiones, la implementación física de las mismas es un punto importante. Las dimensiones pueden contar con múltiples atributos y con jerarquías de varios niveles entre los mismos, lo que hace que se defina si se normalizan o no las tablas de dimensiones diseñadas.

En algunos casos, la dimensión se puede representar en una única tabla donde reside toda la información al nivel más bajo, o también mediante un conjunto de tablas relacionadas que respeten la tercera forma normal.

El primero de los casos es un Esquema Estrella que contiene una tabla de hechos central y un conjunto de tablas de dimensiones relacionadas al hecho. Al normalizar las dimensiones se dice que se transforma al modelo estrella en un copo de nieve, como se muestra en la figura.

---

<sup>1</sup> Fuente Obtenida de: <http://elmejorproyecto2010.blogspot.com/2010/04/data-warehouse-o-almacen-de-datos-y.html>

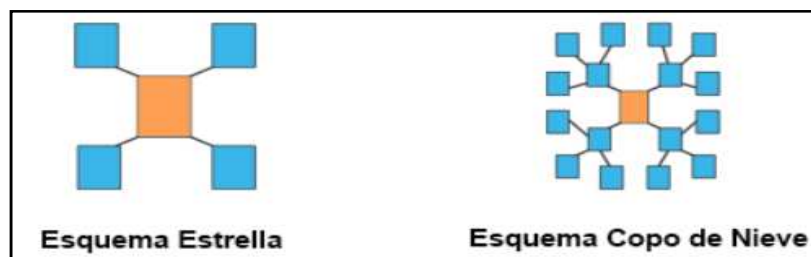


Figura 7. Esquema Estrella y Esquema Copo de Nieve

Las definiciones anteriormente presentadas nos dan una breve introducción al dominio de la creación de los cubos.

### 2.2.1.3 Reportes

Los reportes ofrecen beneficios como distribución de la información rápida, ágil toma de decisiones, toma de decisiones inteligentes, etc. y se pagan por si solos sobre la marcha. Tanto los usuarios finales como el departamento de IT tienen fácil acceso a los datos, no importa donde estos se encuentren. Los reportes traen múltiples analíticos para que usted pueda obtener más respuestas a partir de sus datos y ofrecen adicionalmente una distribución rápida y fácil de los reportes y los resultados.

### 2.2.2 Pentaho

La herramienta que se utilizara para el desarrollo es Pentaho biserver Community Edition 3.5.0.



Pentaho es una herramienta de Inteligencia de Negocios desarrollada bajo la filosofía del software libre para la gestión y toma de decisiones empresariales. Es una plataforma compuesta de diferentes programas que satisfacen los requisitos de BI. Ofreciendo soluciones para la gestión y análisis de la información,

incluyendo el análisis multidimensional OLAP, presentación de informes, minería de datos y creación de cuadros de mando para el usuario.

La plataforma libre de Pentaho para Inteligencia de Negocios cubre muy amplias necesidades de Análisis de los Datos y de los Informes empresariales. Las soluciones de Pentaho están escritas en Java y tienen un ambiente de implementación también basado en Java. Eso hace que Pentaho sea una solución muy flexible para cubrir una amplia gama de necesidades empresariales – tanto las típicas como las sofisticadas y específicas al negocio.



Figura 8. Arquitectura Pentaho

Los módulos de la plataforma Pentaho BI son:

- Reporting - un modulo de los informes ofrece la solución adecuada a las necesidades de los usuarios. Pentaho Reporting es una solución basada en el proyecto JFreeReport y permite generar informes ágil y de gran capacidad. Pentaho Reporting permite la distribución de los resultados del análisis en

múltiples formatos - todos los informes incluyen la opción de imprimir o exportar a formato PDF, XLS, HTML y texto. Los reportes Pentaho permiten también programación de tareas y ejecución automática de informes con una determinada periodicidad.

- **Análisis** - Pentaho Análisis suministra a los usuarios un sistema avanzado de análisis de información. Con uso de las tablas dinámicas (pivot tables, crosstabs), generadas por Mondrian y JPivot, el usuario puede navegar por los datos, ajustando la visión de los datos, los filtros de visualización, añadiendo o quitando los campos de agregación. Los datos pueden ser representados en una forma de SVG o Flash, los dashboards widgets, o también integrados con los sistemas de minería de datos y los portales web (portlets). Además, con el Microsoft Excel Analysis Services, se puede analizar los datos dinámicos en Microsoft Excel (usando la conexión a OLAP server Mondrian).
- **Dashboards** - todos los componentes del modulo Pentaho Reporting y Pentaho Análisis pueden formar parte de un Dashboard. En Pentaho Dashboards es muy fácil incorporar una gran variedad en tipos de gráficos, tablas y velocímetros (dashboard widgets) e integrarlos con los Portlets JSP, en donde podrá visualizar informes, gráficos y análisis OLAP.
- **Data Mining** - análisis en Pentaho se realiza con una herramienta WeKa.
- **Integración de Datos** - se realiza con una herramienta Kettle ETL (Pentaho Data Integration) que permite implementar los procesos ETL. Últimamente Pentaho lanzó una nueva versión - PDI 3.0 – que marcó un gran paso adelante en OSBI ETL y que hizo Pentaho Data Integration una alternativa interesante para las herramientas comerciales.<sup>2</sup>

De los módulos de la plataforma antes mencionados para el desarrollo de la tesis se enfoca en la parte de la Integración de datos, a continuación se especifica con mayor detalle el proceso.

---

<sup>2</sup> Fuente obtenida de : <http://pentaho.almacen-datos.com/>

### 2.2.3 Pentaho PDI (Pentaho Data Integration)

Pentaho Data Integration puede de manera muy simple tomar datos de una fuente de archivos locales y remotos, bases de datos, repositorios y aplicar un procesamiento a dichos datos como filtros, condiciones, cálculos, consultas y almacenar los resultados en un destino como archivos, base de datos o repositorio.

Características de la Herramienta					Java Clove	Java Octopus
¿Forma parte de una plataforma integrada de inteligencia de negocios?	Sí	No	Sí	Sí	No	No
¿La herramienta de extracción posee una interface grafica de uso?	Sí	No	Sí	Sí	No	Sí
¿Soporta diversos tipos de bases de datos?	No	Sí	No	Sí	Sí	Sí
¿Permite cargas desde ficheros excel, xml y planos?	Sí	No	Sí	Sí	Sí	Sí
¿Requiere una fácil instalación de la herramienta?	Sí	Sí	Sí	No	No	No
¿La plataforma posee una herramienta de explotación, herramientas de <i>reporting</i> , herramientas de consultas y análisis?	Sí	No	Sí	Sí	No	No
¿Se encuentra fácilmente consultoras de sistemas para la herramienta?	Sí	Sí	Sí	No	No	No

Tabla 2. Cuadro comparativo de herramientas de extracción<sup>3</sup>

### Programas que forman la herramienta

PDI está formado por un conjunto de herramientas, cada una con un propósito específico.

Spoon: es la herramienta gráfica que nos permite el diseño de las transformaciones y trabajos. Incluye opciones para previsualizar y testear los elementos desarrollados. Es la principal herramienta de trabajo de PDI y con la que construiremos y validaremos nuestros procesos ETL.

<sup>3</sup> Fuente obtenida de:

[http://tesis.pucp.edu.pe/repositorio/bitstream/handle/123456789/931/RODRIGUEZ\\_CABANILLAS\\_KELLER\\_INTELIGENCIA\\_NEGOCIOS\\_ELECTRODOMESTICOS.pdf?sequence=1](http://tesis.pucp.edu.pe/repositorio/bitstream/handle/123456789/931/RODRIGUEZ_CABANILLAS_KELLER_INTELIGENCIA_NEGOCIOS_ELECTRODOMESTICOS.pdf?sequence=1)



Pan: es la herramienta que nos permite la ejecución de las transformaciones diseñadas en spoon (bien desde un fichero o desde el repositorio). Nos permite desde la linea de comandos preparar la ejecución mediante scripts.

Kitchen: similar a Pan, pero para ejecutar los trabajos o jobs.

Carte: es un pequeño servidor web que permite la ejecución remota de transformaciones y jobs.<sup>4</sup>

#### **2.2.4 Pentaho PSW (Pentaho Schema Workbench)**

Pentaho Schema Workbench proporciona en su plataforma BI una solución ROLAP a través de lo que llaman Pentaho Analysis Services. PAS está basado en Mondrian, que es el corazón de este, y en Jpivot, que es la herramienta de análisis de usuario, con el que realizamos la navegación dimensional sobre los cubos desde la plataforma BI y visualizamos los resultados de las consultas. Estas son ejecutadas por Mondrian, que traduce los resultados relacionales a resultados dimensionales, que a su vez son mostrados al usuario en formato Html por Jpivot.

El elemento principal del sistema son los ficheros xml donde se representan los esquemas dimensionales. Para construir estos ficheros xml, se podría utilizar cualquier editor de texto o xml, o bien la herramienta que nos ofrece Pentaho, que se llama Schema Workbench.

Pentaho Schema Workbench es la herramienta gráfica que permite la construcción de los esquemas de Mondrian, y además permite publicarlos al servidor BI para que puedan ser utilizados en los análisis por los usuarios de la plataforma.

En los ficheros de esquema XML, se describen las relaciones entre las dimensiones y medidas del cubo (modelo multidimensional) con las tablas y campos de la base de datos, a nivel relacional. Este mapeo se utiliza para ayudar la traducción de las queries MDX (que es el lenguaje con el que trabaja Mondrian), y

---

<sup>4</sup> Fuente obtenida de: <http://churriwifi.wordpress.com/2010/05/10/16-3-construccion-procesos-etl-utilizando-kettle-pentaho-data-integration/>

para transformar los resultados recibidos de las consultas SQL a un formato dimensional.

### 2.2.5 Pentaho PRD (Pentaho Report Designer)

Pentaho Report Designer es una herramienta de reporting que nos permite crear nuestros propios informes, bien para ejecutarlos directamente o para publicarlos en la plataforma BI y que desde allí puedan ser utilizados por los usuarios. La herramienta es independiente de la plataforma y forma parte del conjunto de herramientas de la suite de Pentaho.

Pentaho provee acceso a fuentes de datos relacionales, OLAP o basadas en XML, además de ofrecer varios formatos de salida como: PDF, HTML, Excel o archivos planos.



Figura 9. Report Designer

### **2.3 Identificación de Variables**

Variables independientes:

Falta de organización de la información, Fuentes pobres de consulta, Diversas fuentes de información, Control limitado por parte de los directores de carrera para obtener la información requerida, Tiempo de respuesta.

Variables dependientes:

Mejor registro de información, Fuente de datos centralizados y autónomos, Unificación de la Información, Facilidad de acceso a la información para el control y toma de decisiones de la carrera, Mejor tiempo de respuesta.

### **2.4 Hipótesis**

La realización del proyecto Análisis, Diseño e Implementación de un Datamart Académico usando Tecnología de BI para la Facultad de Ingeniería, Ciencias Físicas y Matemáticas, permitirá mejorar el control y análisis de la información de la Facultad de modo que los Directores de Carrera puedan tener acceso a la información cuando lo requieran y puedan realizar la toma de decisiones estratégicas.

## **CAPITULO III**

### **3 MARCO METODOLOGICO**

#### **3.1 Diseño de la Investigación**

El presente proyecto de tesis desea implementar un modelo de control para la toma de decisiones gerenciales tomando como aplicación la creación de Reportes e Indicadores de Control, el cual ayudará con la Planificación estratégica para la toma de decisiones para el bien de la Facultad; así como también se emplearán mapas estratégicos que nos permitan canalizar las metas de la organización en objetivos concretos.

La investigación a realizar para este proyecto es la Investigación de Campo, que va a permitir mediante la manipulación de datos obtener información que sirva de apoyo a la Facultad, permitirá conocer en qué condiciones se encuentra el manejo de la información, así como también poder determinar las mejores prácticas que se podrían aplicar para el proceso técnico de la facultad.

Logrando que el conocimiento de los procesos en la Facultad sea transmitido más a fondo al desarrollador del proyecto, de modo que se pueda manejar los datos con más seguridad y poder soportarse en diseños exploratorios, descriptivos y experimentales, creando una situación de control en la manipulación de la información, como lo requieran.

#### **3.2 Diseño Experimental.**

En la realización del proyecto de tesis, para el diseño experimental se analiza como las variables independientes de la investigación, afectan a los procesos en la facultad que este caso es la falta de la organización de la información, fuentes pobres de consulta, que provocan problemas en la toma ágil de decisiones, control y organización de la información en la facultad.

Razón por la cual se automatizo el control de la información, centralizando la misma en Data Marts y creando Reportes e Indicadores Estratégicos que facilite el

análisis y consulta a los distintos usuarios de la aplicación, de acuerdo a su perfil la información se muestra en indicadores estratégicos.

### **3.3 Plan de Recolección de Datos.**

Una vez planteada la investigación y realizadas todas las tareas que permitan la puesta en marcha de la investigación, se efectúa la recolección de datos e información, con el fin de dar respuesta al problema planteado como propuesta de tesis.

Las técnicas que se utilizo en la recolección de datos son:

- La observación del manejo de la información en la facultad.
- Las entrevistas con los stakeholders del aplicativo que se realizo como son las autoridades de la Facultad y potenciales usuarios del aplicativo, entre otros.
- Recopilación documental de toda la información que se maneja para la Facultad y que se encuentra en distintos formatos.
- Análisis de contenido una vez que se tenga la información académica de la Facultad, para determinar el tratamiento de la información.

Constituye una fase completamente mecánica, la cual puede llevarse a cabo de forma manual o computarizada. Todo depende del tipo de método que se determino utilizar en las distintas etapas del desarrollo de la aplicación.

### **3.4 Plan de Procesamiento de Datos**

Una vez obtenida la información recopilada para realizar el proyecto de tesis, el procesamiento de datos para la investigación a realizar consiste en:

Utilizar la información consolidada en un Data Mart en el cual se almacenan los datos académicos de los estudiantes de la Facultad y luego proceder con la creación de los Reportes e Indicadores Estratégicos para la toma de decisiones.

De modo que permita facilitar el registro y consulta de los datos para los usuarios del sistema y poder ofrecer la información requerida por las autoridades de la Facultad para el control de los mismos.

### **3.5 Metodología de Desarrollo**

La Metodología de desarrollo a utilizarse es El Proceso Unificado Racional (Rational Unified Process en inglés, habitualmente resumido como RUP), que divide la implementación en fases que son:

- **Planificación:** Analizar y entender las necesidades del negocio para el cual se está desarrollando el software.
- **Análisis de Requerimientos:** Proveer una base para estimar los costos y tiempo de desarrollo del sistema.
- **Elaboración:** Trasladar los requisitos analizados anteriormente a un sistema automatizado y desarrollar una arquitectura para el sistema.
- **Construcción:** Crear software que se ajuste a la arquitectura diseñada y que tenga el comportamiento deseado.
- **Pruebas:** Asegurarse de que el comportamiento requerido es correcto y que todo lo solicitado está presente.
- **Transición:** Producir distribuciones del producto y distribuirlo a los usuarios.

## CAPITULO IV

### 4 DISEÑO DE LA SOLUCIÓN

#### 4.1 Descripción de la Arquitectura

Para tener una visión general del sistema y tomando en cuenta la plataforma elegida para la implementación de la solución Pentaho, a continuación se explica la arquitectura común utilizada en los sistemas de Data Mart, detallando cada uno de los procesos o sub-sistemas que conforman el proyecto de tesis.

Se tiene seis procesos o sub-sistemas en los cuales se encuentra estructurado el proyecto de tesis:

- Fuente de datos
- Extracción, transformación y carga
- Cubos de Información
- Presentación (Reportes e Indicadores)
- Seguridad
- Administración

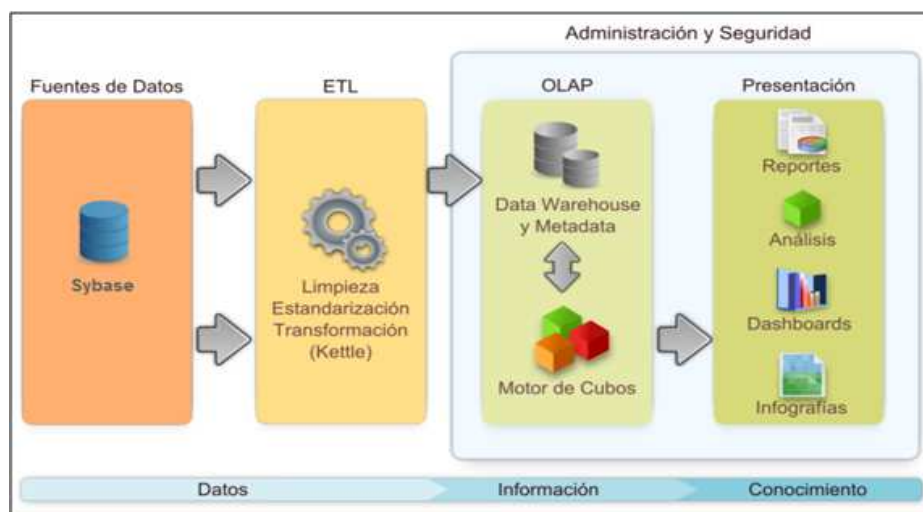


Figura 10. Arquitectura de la Solución

El sub-sistema corresponde a las fuentes de datos que se utilizaron en la obtención de los datos que alimenta el sistema. En este proyecto se utilizaron como fuente de datos una base relacional que es Sybase y archivos planos.

El módulo del ETL es la sección donde se agrupan una serie de procesos que llevan a cabo tareas relacionadas con la extracción, manipulación, control, integración, limpieza de datos, carga y actualización de los Data Marts, todas las tareas que se hagan desde que se toman los datos de la fuente en este caso la base relacional y archivos planos, hasta que se carguen en el sistema para su utilización en la construcción de los cubos o reportes.

En esta parte del sistema se mantiene los datos obtenidos en una base de datos temporal que se encuentra en Postgresql, que es usada para ejecutar todos los procesos, tareas antes mencionadas.

El sistema OLAP es el núcleo del sistema que es el repositorio central de la información donde se almacenan los datos que se están utilizando. En el Data Mart se almacenan los datos operacionales en estructuras multidimensionales que optimizan el acceso para las consultas y que son flexibles, además de contener toda la metadata de la información almacenada que ofrece información descriptiva sobre el contexto, la calidad, condición y características de los datos. En esta área se incluye el motor de cubos multidimensional que es el encargado de ejecutar las consultas realizadas por los componentes externos.

El módulo de presentación es el área correspondiente a la interacción con el usuario, cuya funcionalidad es mostrar los datos almacenados de forma útil y transparente a través de las distintas herramientas. Este sistema se comunica directamente con el servidor de cubos a través de consultas, las cuales retornan la información requerida donde ésta es transformada y presentada para la visualización final. Los reportes requeridos en el proyecto se encuentran en esta área.

En módulo de Seguridad se encuentran definidas las restricciones de acceso a los objetos de la plataforma y a los diferentes recursos.



Y finalmente, en el módulo de administración se encuentran las herramientas administrativas de la plataforma. Gestión de usuarios, administración de conexiones de fuentes de datos, herramientas de limpieza de los diferentes cachés y el sistema de archivos interno del Data Mart se encuentran en esta área.

#### **4.2 Análisis de las fuentes de datos**

Los datos que se utilizaron para alimentar el Data Mart corresponden al la fuente de datos del Sistema de la facultad (Fing). Este sistema tiene información relacionada a los docentes y estudiantes de la facultad.

La base de datos no tiene suficientes restricciones referenciales ni método de verificación de los datos que se ingresan, pudiendo así ingresar información no solo incorrecta, sino también en diferentes formatos con significados no siempre interpretados de la misma forma.

Los reportes e informes generados a petición de los directores de carrera son generados de forma manual y la mayoría de estos entregados como archivos planos de texto.

Este sistema es la principal fuentes de información que se utilizo a lo largo del proyecto. Puntualmente, es el proveedor de los datos a partir del cual se generaron los reportes e indicadores. Es por ello que se realizó un análisis exhaustivo de la base de datos para diferenciar cuales son las entidades relevantes para la solución del problema.

A partir del análisis de las vistas entregadas por la persona encargada de la fuente se procedió a analizar los requerimientos y obtener información de cuáles eran los campos necesarios para los reportes e indicadores solicitados, se lograron identificar cuáles eran los campos que formarían parte de la solución.

A continuación se detallan dichas vistas:

Se tiene una vista Carga\_Horaria, esta vista tienen datos de los docentes, materia que imparte, la carga horaria que tienen, categoría, dedicación y periodos en los que han impartido conocimientos en la Facultad.

La segunda vista es Estudiante\_Indicadores, esta vista tiene datos de los estudiantes, información personal, de que institución vienen, fecha de inscripción, matrícula, detalla el record académico de los estudiantes de la facultad de todas las carreras.

A continuación se muestran las tablas involucradas como fuente de datos del sistema.

Carga_Horaria	Estudiante_Indicadores
hora_hora_inicio	fac_nombre
prof_codigo	esc_nombre
conxprof_fecha_inico	per_nombre
con_nombramiento	est_codigo
con_categoria	est_nombre
cur_codigo	est_sexo
mat_codigo	est_fecha_nacimiento
esc_codigo	est_nota_grado_colegio
per_codigo	pai_nombre
esp_nombre	ciu_nombre
per_nombre	col_nombre
hora_hora_final	espe_nombre
hora_codigo_dia	nac_nombre
conxprof_fecha_final	est_estado_civil
con_dedicación	ins_estado
aula_nombre	per_codigo
matxprof_paralelo	ins_num_matricula_curso
fac_codigo	fac_codigo
esp_codigo	esc_codigo
nombre_profesor	esp_nombre
prof_cedula_id	cur_nombre
mat_nombre	est_apellido
	est_cedula
	est_modalidad
	est_estado
	pai_codigo
	ciu_codigo
	col_codigo
	col_esp_codigo
	nac_codigo
	est_edad
	est_estado_trabajo
	ins_fecha
	cur_codigo
	esp_codigo
	pen_codigo

Figura 11. Vistas del Sistema Fing.

### 4.3 Calidad de Datos

Al analizar las fuentes de información se detectaron problemas con la calidad de los datos contenidos en las mismas, lo cual interfirió en la implementación del

Data Mart. Fue necesario solucionar estos inconvenientes para poder avanzar con la carga de datos. La eficiencia en la toma de decisiones depende directamente de la calidad de los datos analizados, de modo que toda mejora en la exactitud de los mismos genera mejoras sustanciales en las decisiones tomadas. De la misma forma en que se mejora los niveles de exactitud de los datos se gana en credibilidad sobre la herramienta, haciendo que se fomente el uso de la misma por parte de los usuarios.

Los inconvenientes en los datos se manejaron en dos etapas. La primera etapa tuvo como objetivo principal realizar un análisis que permitiera descubrir posibles valores inválidos, faltantes o mal representados, a los cuales se les aplicó acciones correctivas manuales o automatizadas a través de los procesos de carga. Esto permitió que los datos que sean integrados en el Data Mart coincidan a lo que representa la realidad. La segunda etapa consistió en notificar cuáles son los problemas que no podrían ser solucionados por parte de los desarrolladores del proyecto a las personas encargadas del sistema fuente para que puedan validar por su cuenta las observaciones encontradas y mejorarlos en términos de calidad de la información.

Con este análisis se trató de lograr la mejor correspondencia entre la estructura de la base de datos de origen y las tablas temporales del Data Mart, a la vez se buscó recuperar la mayor cantidad posible de datos para proveer un bloque de información robusto y confiable para la toma de decisiones.

En la primera etapa, se analizó tres aspectos que permitieron realizar la búsqueda de inconsistencias fueron separados en los siguientes puntos:

Análisis de meta data: Mediante la comparación de los valores, las definiciones relacionales en las cuales los mismos estaban almacenados y los formatos definidos.

Análisis de contenido: A través del análisis de las reglas de la institución y las propiedades de las columnas de los atributos.

Análisis de relaciones: Mediante verificaciones de integridad y búsqueda de datos redundantes.

Una vez realizado este análisis se diferencia un conjunto importante de problemas. Los mismos se numeran a continuación:

- Errores de integridad entre valores de diferentes tablas (ejemplo: la edad de una persona es distinta a la diferencia entre la fecha de nacimiento y la fecha de realizado el registro).
- Problemas para identificar el pensum. (ejemplo: en los últimos años a existido un constante cambio de pensum en la facultad).
- Valores fuera de rangos posibles (ej: edades y fechas).
- Valores de columnas con errores tipográficos (ejemplo: se encontró fechas mal ingresadas, ya que en vez de ingresar un año de nacimiento con el año 2000, se lo ingresaba con el año 2050).
- Datos mal ingresados (ejemplo: estudiantes de sexo femenino estaban categorizados como masculino y viceversa).

#### **4.4 Frecuencia de Carga**

En particular en este proyecto el problema de la frecuencia de carga está acotado, dado que tomando en cuenta las vistas utilizadas para la implementación, en el peor de los casos los datos más actuales tendrían un mes de antigüedad dentro del sistema.

#### **4.5 Modelado Multidimensional**

En esta sección se denotan cada uno de los elementos multidimensionales que forman parte de la solución. Para el modelado de la solución se tomo como base la lista de indicadores proporcionado por los directores de las distintas carreras de la facultad, los cuales hicieron un análisis previo para la creación de estos requerimientos, el cual es la base de los requerimientos usados para la construcción y modelamiento del proyecto.

Las dimensiones que se derivan de los requerimientos son las que se describen a continuación:

### **Dimensión Tiempo**

Tiempo es la dimensión que determina a que mes, semestre, año, etc. pertenecen los datos cargados dentro del Data Mart y sobre los cuales se realiza las consultas en función del tiempo.

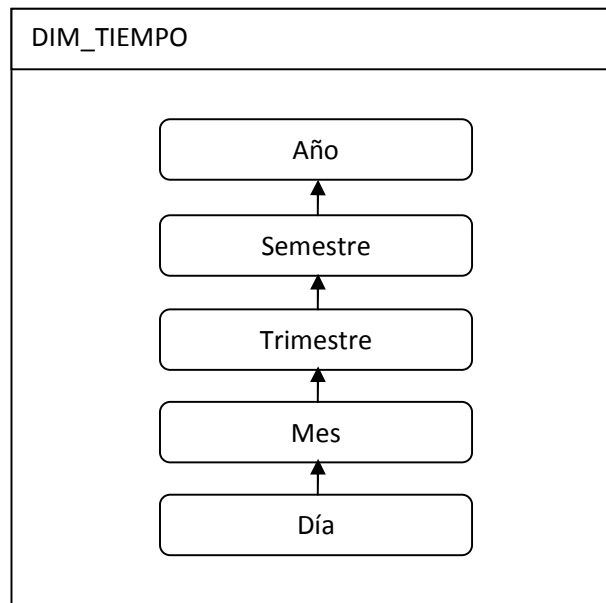


Figura 12. Dimensión Tiempo

### **Dimensión Periodo Lectivo**

La dimensión Periodo Lectivo se utiliza para el análisis de la mayoría de los indicadores ya que en esta dimensión se cargan los datos del código y periodo lectivo de todo el record académico del estudiante.

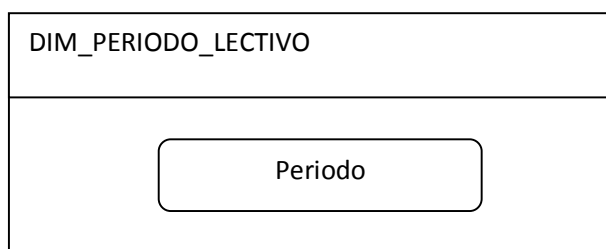


Figura 13. Dimensión Periodo Lectivo

### **Dimensión Categoría Docente**

La dimensión Categoría Docente se utiliza para el análisis de los indicadores de docentes se carga con los datos del código y descripción de la categoría de los docentes.

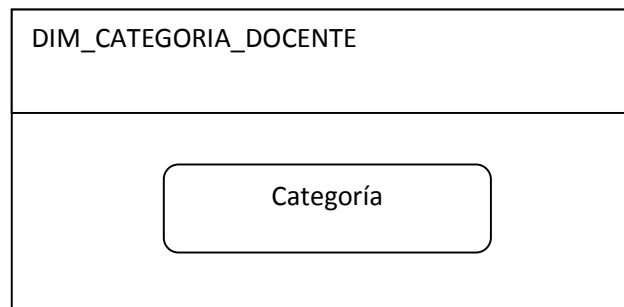


Figura 14. Dimensión Categoría Docente

### **Dimensión Dedicación Docente**

La dimensión Dedicación Docente se utiliza para el análisis de los indicadores de docentes se carga con los datos del código y descripción de la dedicación de los docentes.

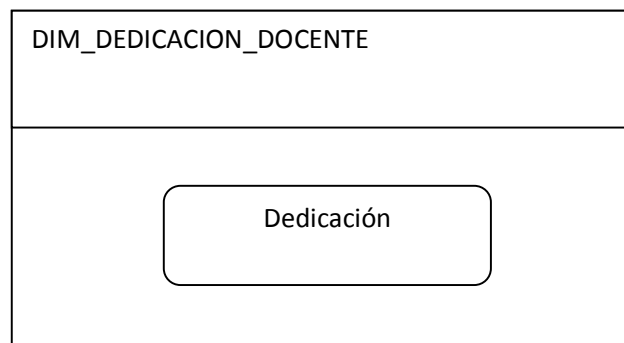


Figura 15. Dimensión Dedicación Docente

### **Dimensión Facultad**

La dimensión Facultad se utiliza para el análisis de todos los indicadores tanto de docentes como estudiantes se carga con los datos del código y descripción de cada una de las carreras que existe en la Facultad de Ingeniería. Se tiene una

granularidad hasta un nivel más bajo que es la materia que toma el estudiante en cada periodo lectivo.

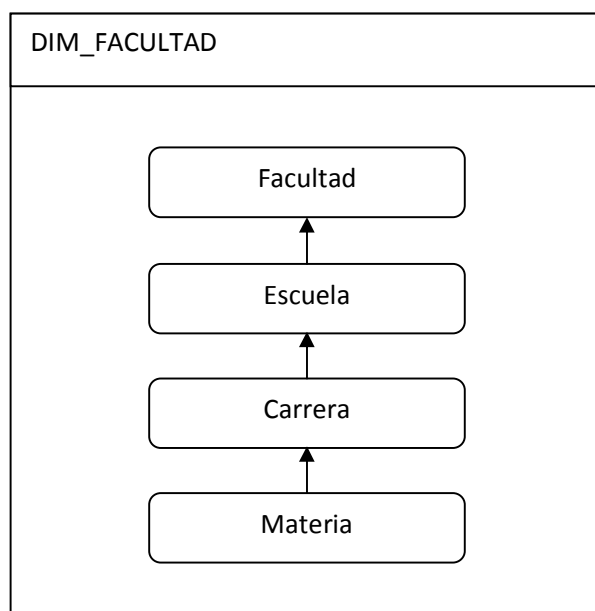


Figura 16. Dimensión Facultad

### Dimensión Especialidad

La dimensión Especialidad se utiliza para el análisis de los indicadores de docentes se carga con los datos del código y descripción de la facultad, escuela, carrera en la que los docentes imparten clases.

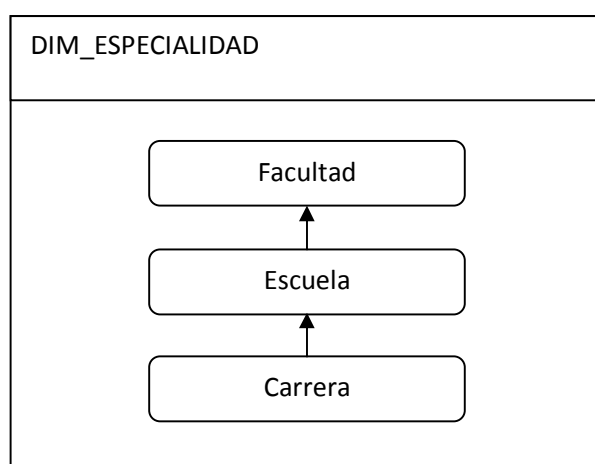


Figura 17. Dimensión Especialidad

### Dimensión Estado Estudiante

La dimensión Estado Estudiante se utiliza para el análisis de los indicadores de estudiantes se carga con los datos del código y descripción del estado del estudiantes, esta dimensión dice si el estudiante es activo, egresado, supera el número de matrícula o es cambio de malla.

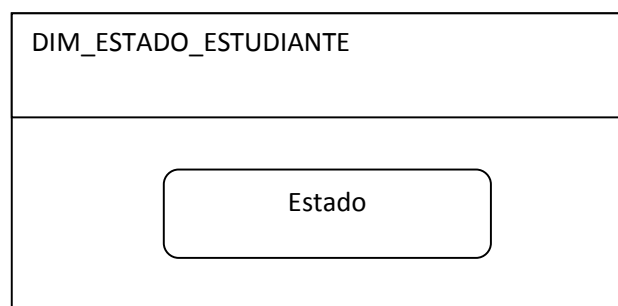


Figura 18. Dimensión Estado Estudiante

### Dimensión Profesor

La dimensión Profesor se utiliza para el análisis de los indicadores de docentes se carga con los datos del código y descripción del docente, esta dimensión muestra los datos básicos del docente .

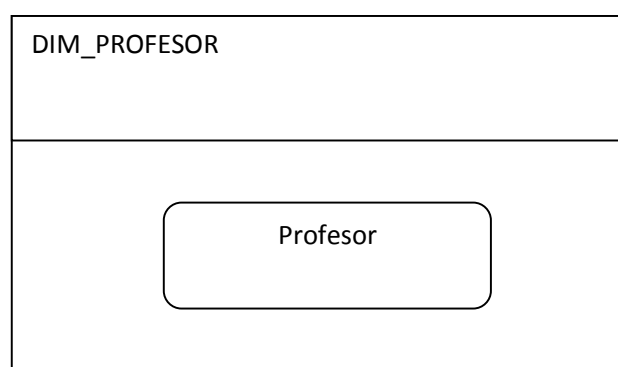


Figura 19. Dimensión Profesor

### Dimensión Estudiante

La dimensión Estudiante se utiliza para el análisis de los indicadores de estudiantes se carga con los datos del código y descripción de los datos básicos del



estudiante, esta dimensión contiene datos del estudiante como nombres, apellido, sexo, país, de que colegio viene, etc.

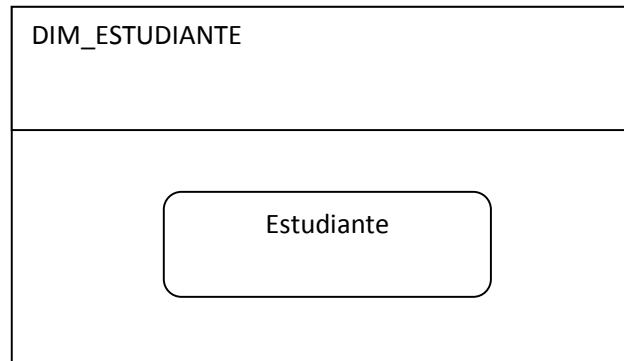


Figura 20. Dimensión Estudiante

### **Dimensión Estado Curso Estudiante**

La dimensión Estudiante Curso Estudiante se utiliza para el análisis de los indicadores de estudiantes se carga con los datos del código y descripción del estado del estudiante en el periodo lectivo, esta dimensión contiene datos del estudiante como si está inscrito, matriculado o anulo la matrícula.

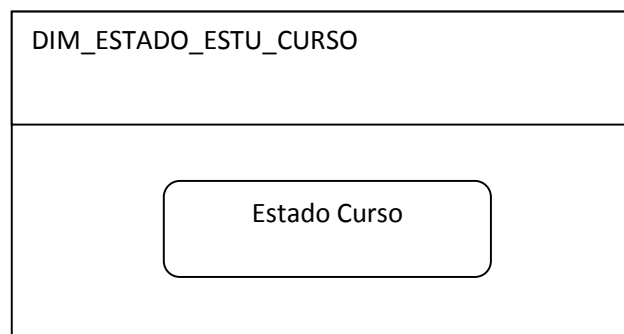


Figura 21. Dimensión Estado Curso Estudiante

### **Dimensión Curso**

La dimensión Curso se utiliza para el análisis de los indicadores de estudiantes y docentes se carga con los datos del código y descripción del curso en el que se encuentra el estudiante o curso donde da clases el docente.

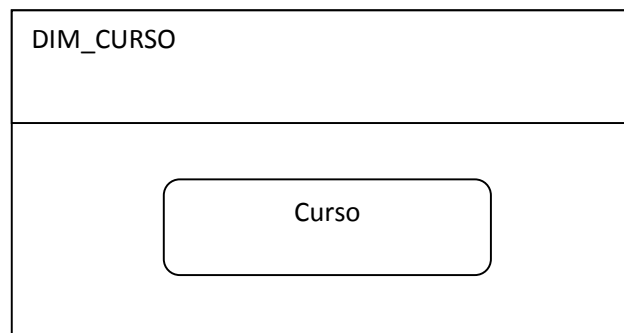


Figura 22. Dimensión Curso

#### 4.5.1 Esquema Multidimensional

En las figuras a continuación se muestran los esquemas multidimensionales que se definieron para la construcción de los cubos de información.

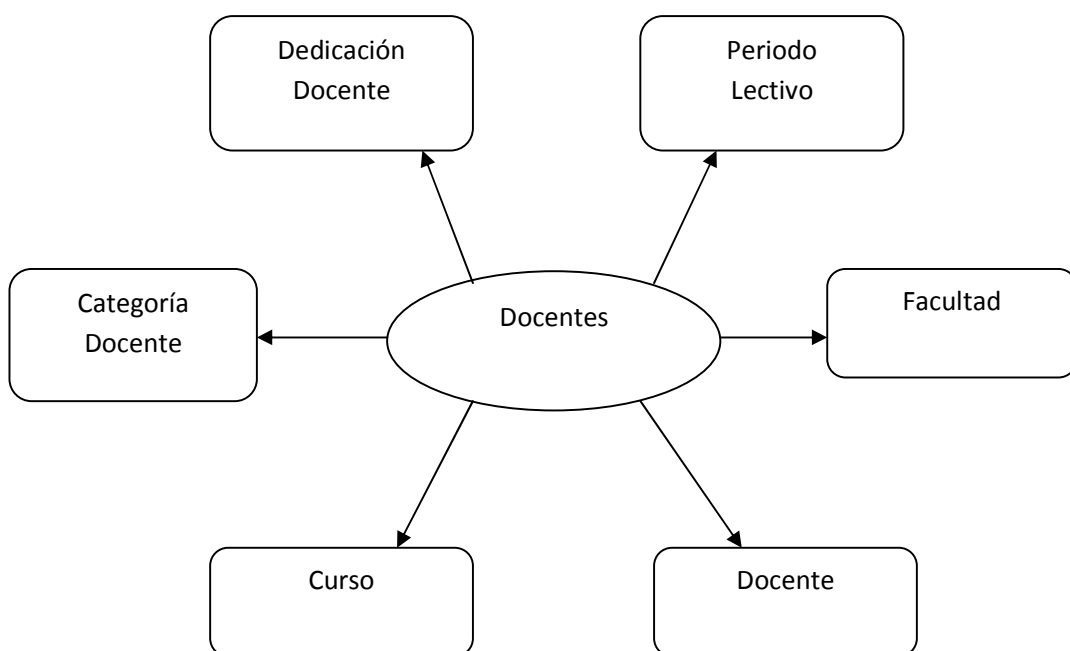


Figura 23. Relación dimensional para el cubo de Docentes

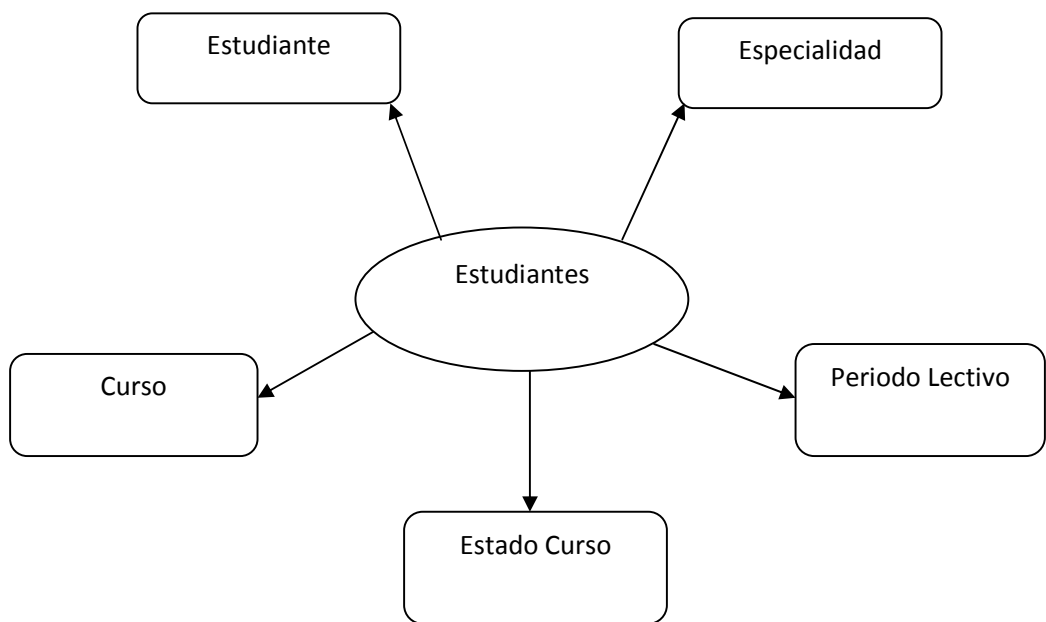


Figura 24. Relación dimensional para el cubo de Estudiantes

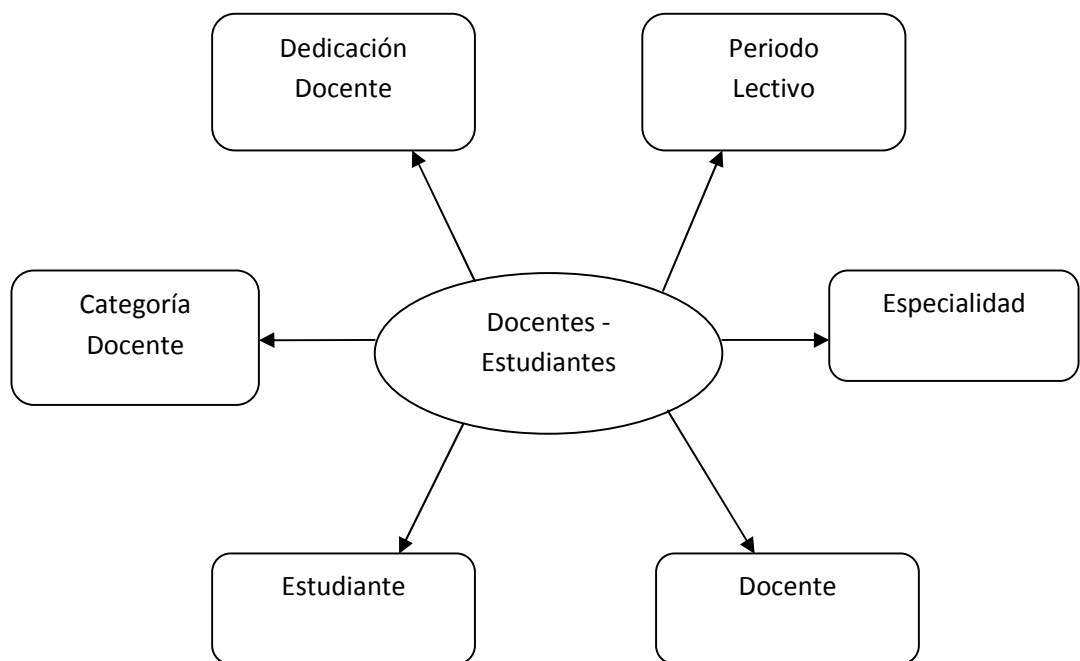


Figura 25. Relación dimensional para el cubo de Docentes – Estudiantes

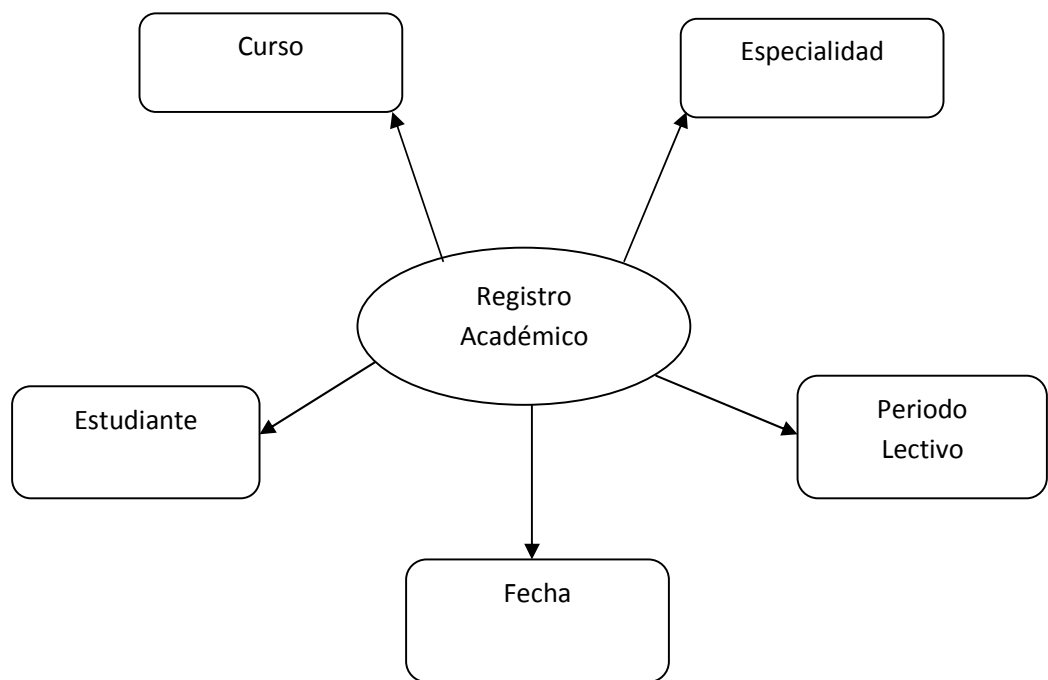


Figura 26. Relación dimensional para el cubo de Registro Académico

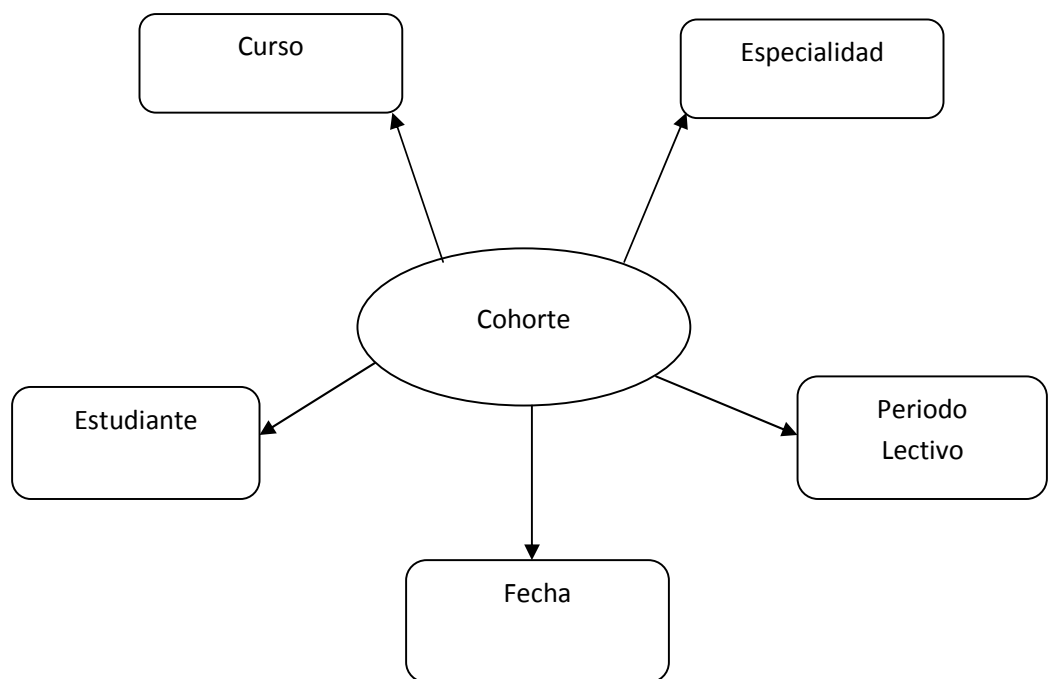


Figura 27. Relación dimensional para el cubo de la Cohorte

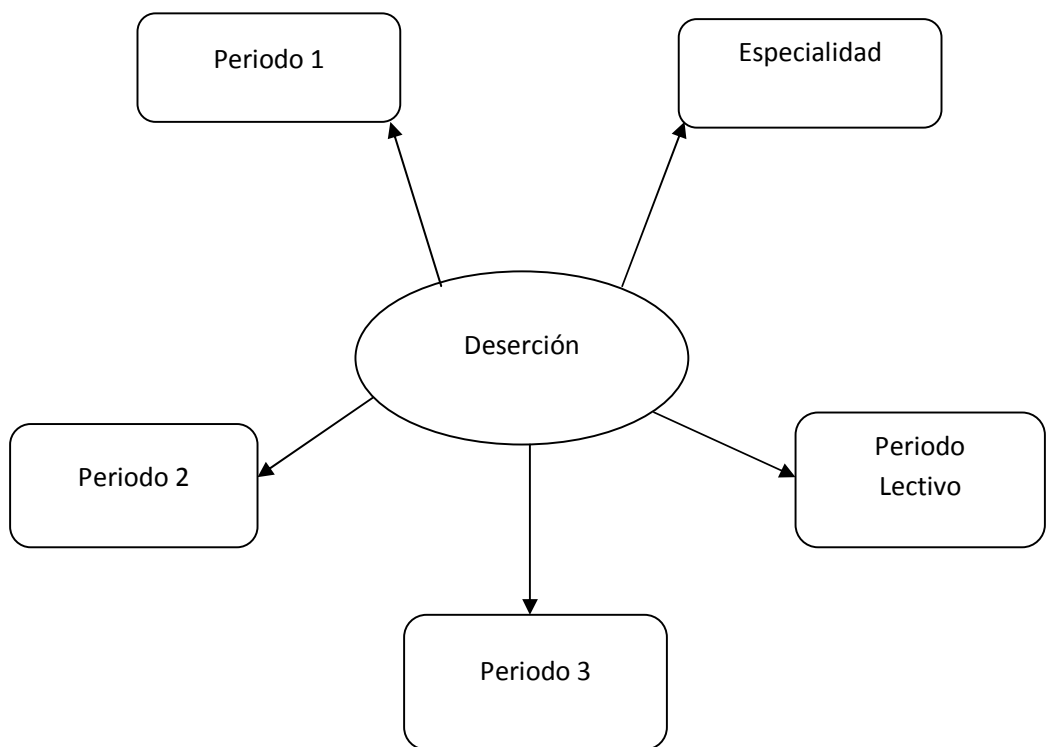


Figura 28. Relación dimensional para el cubo de Deserción

#### 4.6 Diseño relacional de la base que soporta a los cubos

Las figuras a continuación representan los diagramas relacionales de las estructuras relacionales usadas para la creación del modelo multidimensional usado para el manejo de la construcción de los cubos para los indicadores y reportes.

Éste se deriva de las dimensiones y del esquema multidimensional. Se utilizó un esquema estrella para el diseño en lugar de un esquema copo de nieve, ganando así simplicidad en el diseño y velocidad de acceso para obtener tener las distintas jerarquías.

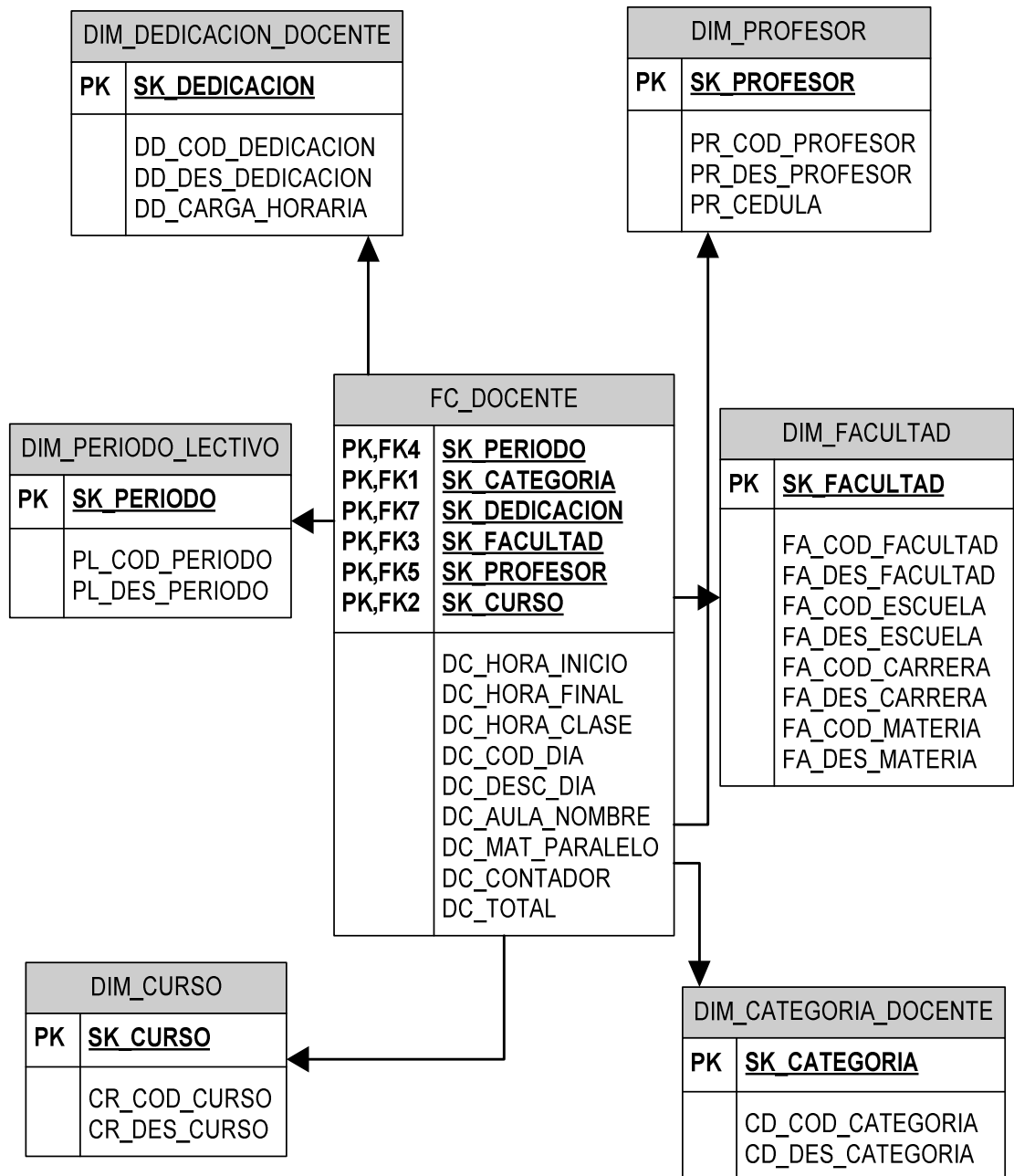


Figura 29. Modelo Multidimensional en base a tablas relacionales para el cubo de DOCENTES

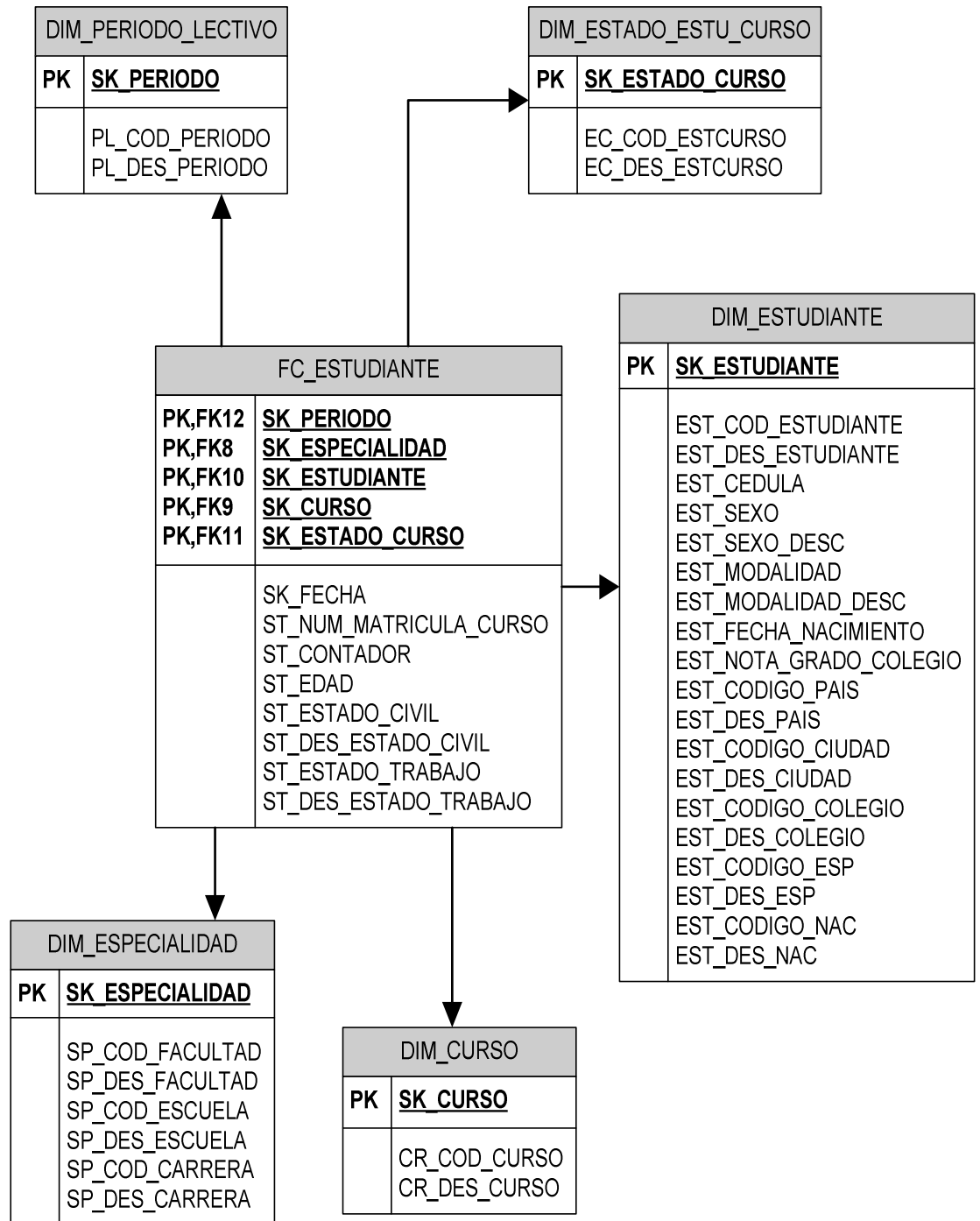
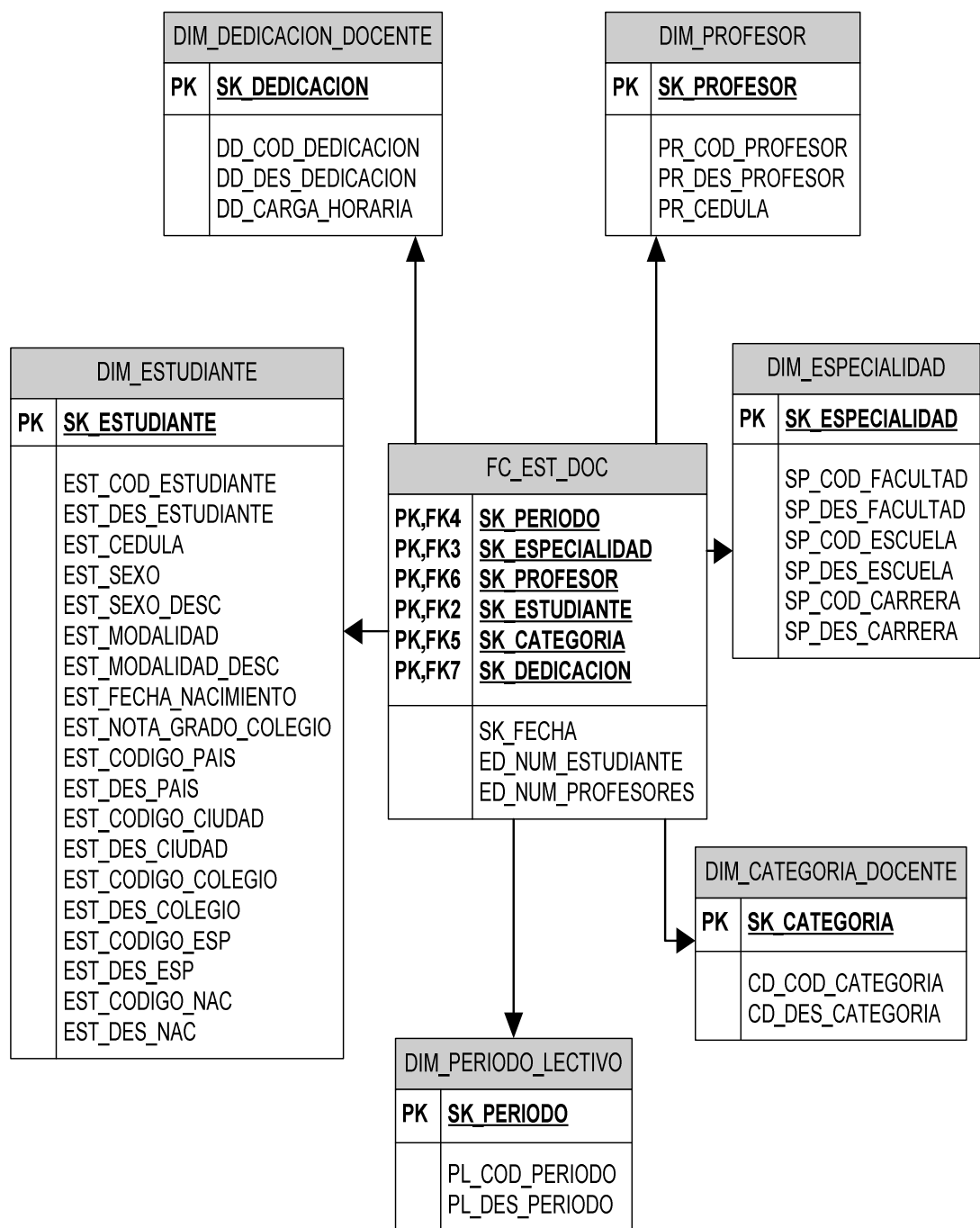
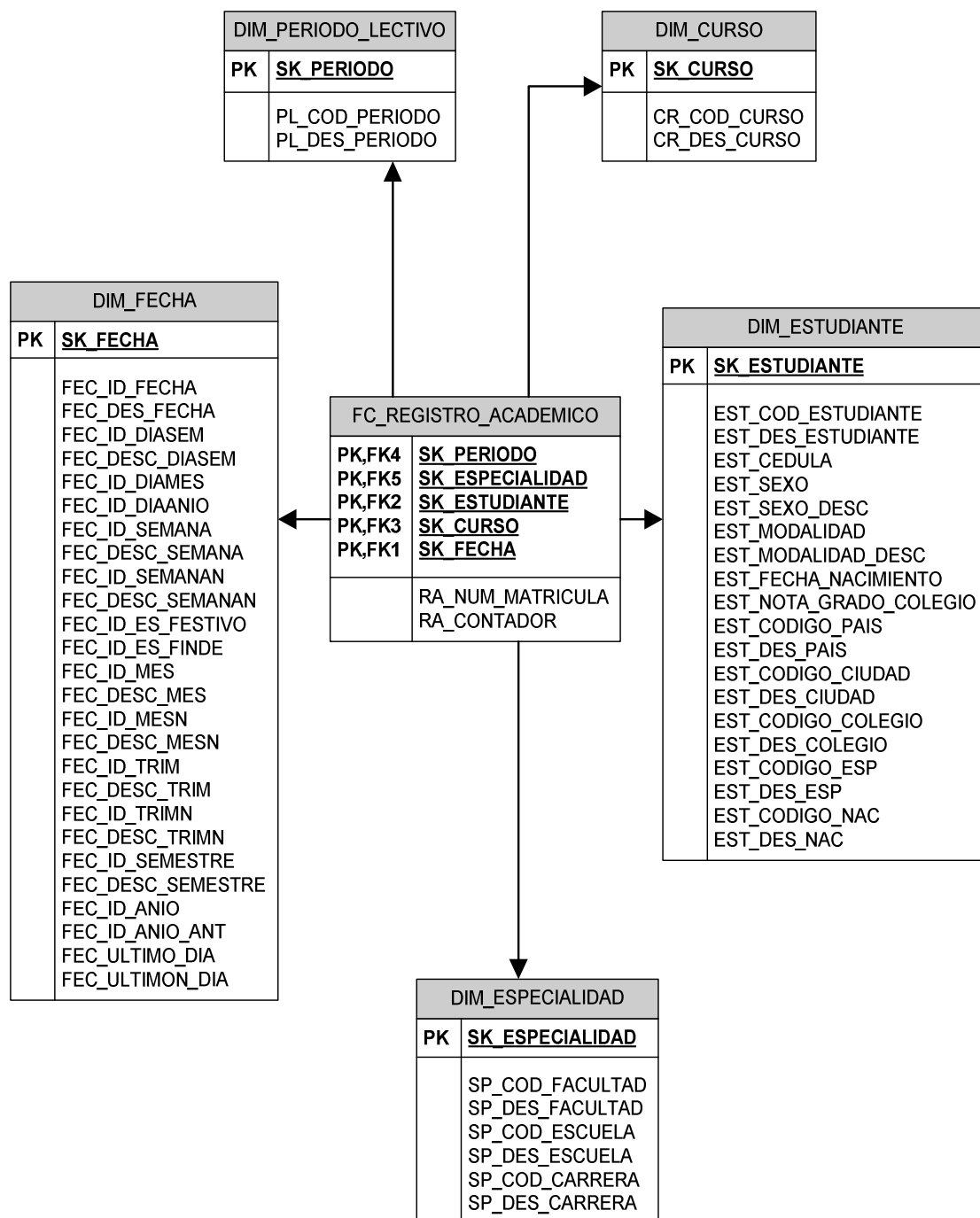


Figura 30. Modelo Multidimensional en base a tablas relacionales para el cubo de ESTUDIANTES



**Figura 31.** Modelo Multidimensional en base a tablas relacionales para el cubo de DOCENTE - ESTUDIANTE





**Figura 32.** Modelo Multidimensional en base a tablas relacionales para el cubo de REGISTRO ACADEMICO

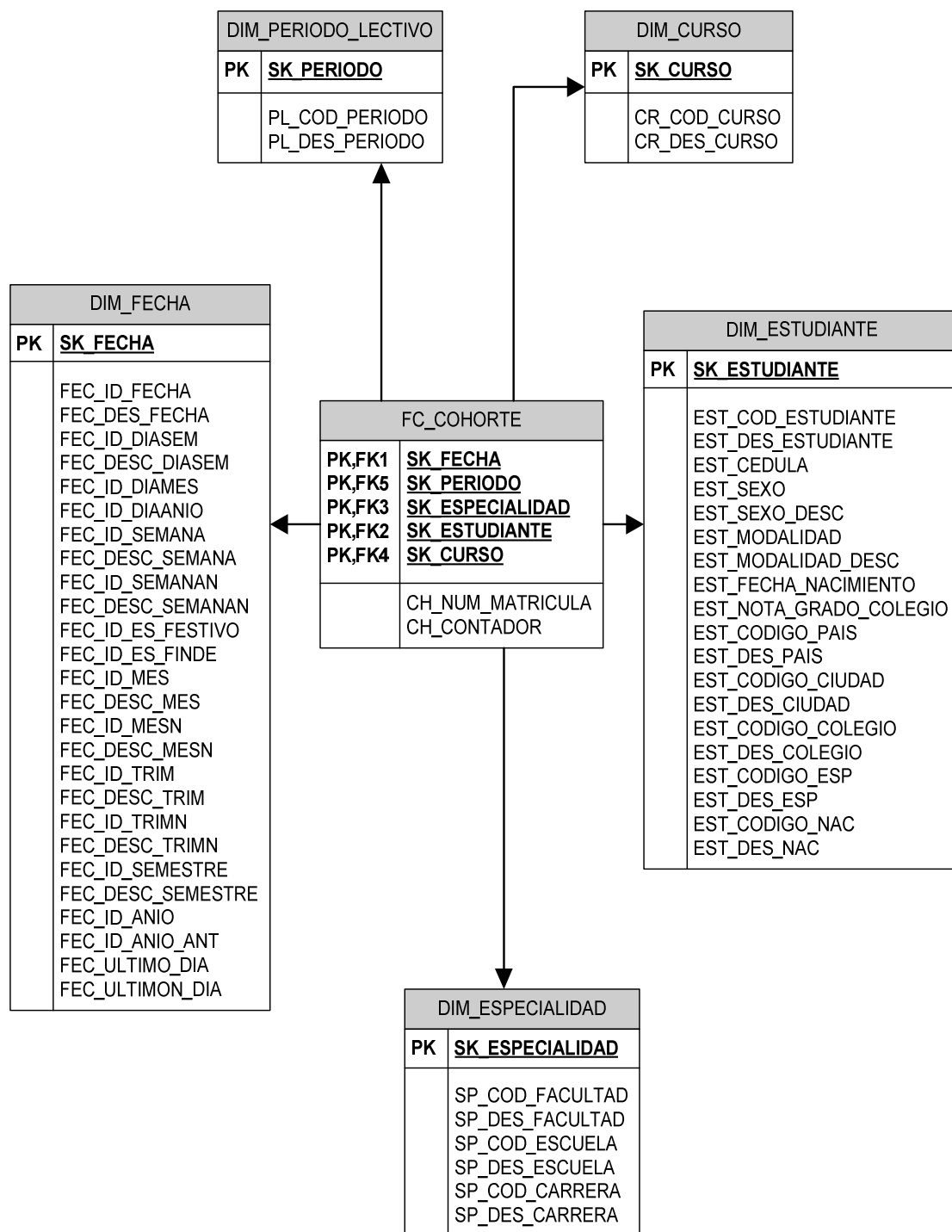


Figura 33. Modelo Multidimensional en base a tablas relacionales para el cubo de la COHORTE

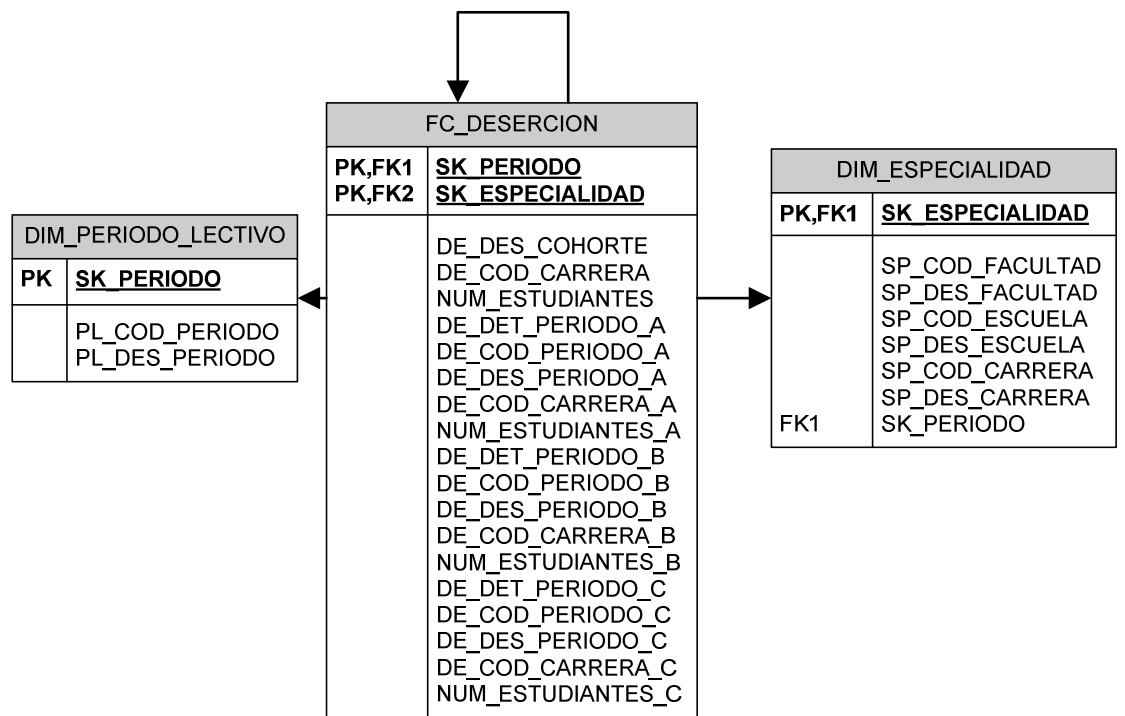


Figura 34. Modelo Multidimensional en base a tablas relacionales para el cubo de la DESERCIÓN

## CAPITULO V

### 5 IMPLEMENTACION DEL PROTOTIPO

En este capítulo se detalla el proceso de implementación de la solución diseñada.

#### 5.1 Componentes de la arquitectura de la suite de Pentaho

La arquitectura funcional y los diferentes componentes que forman parte de la suite de Pentaho, se puede separar en cuatro bloques principales que son:

- Orígenes de datos: Es donde se encuentran los sistemas desde los cuales se extrae la información.
- Integración de datos: Es la capa en donde se encuentran la herramienta de ETL (Data Integration) y la de creación de metadata de cubos (Schema Workbench).
- Plataforma de Business Intelligence (BI): Es el conjunto de herramientas que permiten la administración y ejecución de los artefactos creados para realizar el análisis de los datos. En esta capa se encuentra el repositorio de archivos, la lógica de negocios, los sistemas administrativos de la plataforma y los componentes que gestionan la seguridad sobre todos los artefactos.
- Capa de presentación: Es donde se visualizan los resultados de las ejecuciones de los distintos artefactos creados para realizar el análisis.

Para el desarrollo del proyecto de tesis se utilizaron la mayoría de los componentes propuestos por la plataforma. Entre los más relevantes se encuentran el componente de ETL para la extracción transformación y carga de los datos y todos los componentes que conforman la capa que contiene la plataforma de BI. Sobre la capa de presentación de los datos se implementaron sobre el análisis, reportes e indicadores. Dentro de la capa de Integración de datos otra herramienta utilizada para crear la metadata de los cubos.

El paquete más importante, dado que es utilizado en todas las capas de la plataforma son los cubos OLAP, que son las estructuras multidimensionales que

guardan la información dentro de nuestro sistema y que permiten realizar operaciones sobre la misma.

## 5.2 Extracción, Transformación y Carga

En esta sección se explica los pasos seguidos para realizar la extracción, transformación y carga de datos dentro de los distintos Data Mart desde las fuentes de datos externas a la plataforma. Se referirá al proceso de extracción, transformación y carga de aquí en adelante como ETL.

### 5.2.1 Introducción

Para el proceso ETL se utiliza la herramienta Pentaho Data Integration (PDI) de la plataforma Pentaho, también conocida como Kettle.

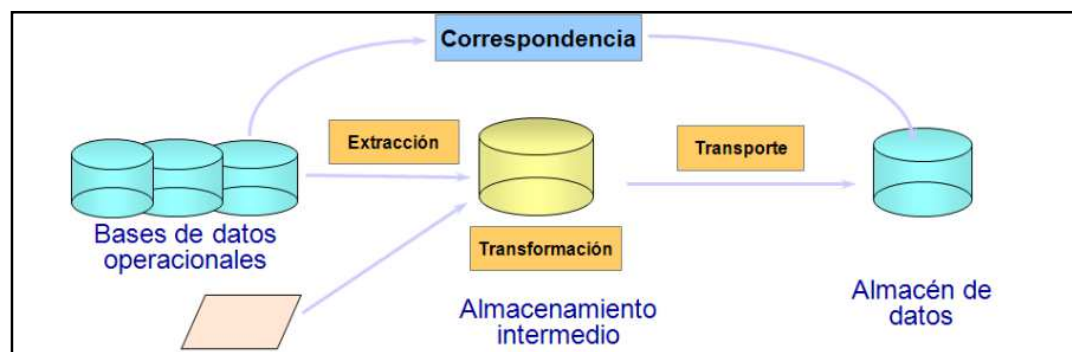


Figura 35. Proceso ETL

PDI cuenta con varias aplicaciones para la manipulación de datos, entre las que se tienen Spoon, Pan y Kitchen. Por medio de Spoon, se puede acceder a la interfaz gráfica que permite diseñar los elementos principales de un proceso de ETL, los cuales son: trabajos y transformaciones.

Pan es una aplicación que interpreta y ejecuta transformaciones diseñadas con Spoon.

Kitchen es una aplicación que interpreta y ejecuta trabajos diseñados con Spoon.

Mediante estas 3 herramientas se cubren todos los requerimientos de carga del Data Mart.

A continuación se detalla los procesos ETL creados con las herramientas para el proyecto.

## 5.2.2 Procesos ETL

### 5.2.2.1 Procesos ETL para carga de datos al STAGE

Este proceso gestiona las transformaciones y trabajos intermedios a ejecutarse y realizan la carga ordenada de los datos requeridos hacia las tablas del STAGE que serán fuente para la carga de datos de las dimensiones y hechos a utilizarse para la creación de los cubos de información.

A continuación se detallan los procesos creados:

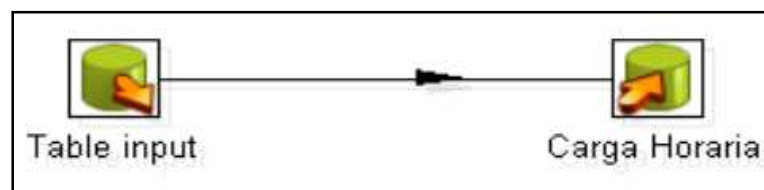


Figura 36. Transformación STG\_CARGA\_HORARIA

Esta transformación crea una tabla en la base de datos STG\_DWH, llamada STG\_CARGA\_HORARIA contiene toda la información de la vista CARGA\_HORARIA de la fuente de datos original.

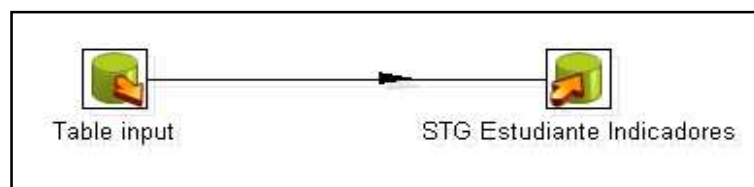
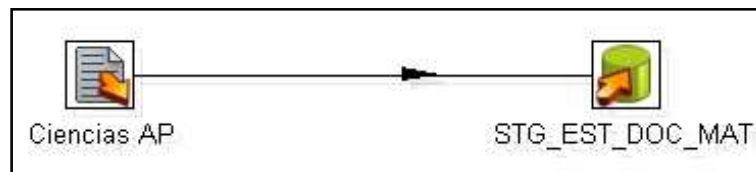


Figura 37. Transformación STG\_ESTUDIANTE\_INDICADORES

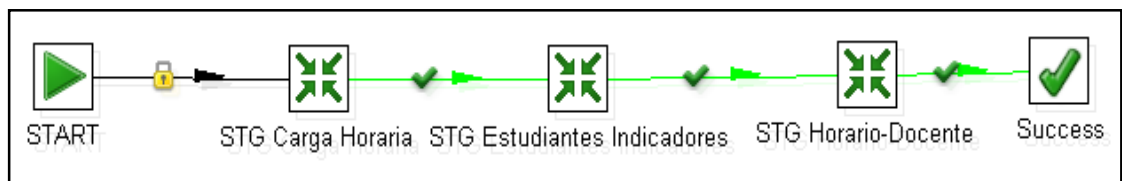
Esta transformación crea una tabla en la base de datos STG\_DWH, llamada STG\_ESTUDIANTE\_INDICADORES contiene toda la información de la vista ESTUDIANTE\_INDICADORES de la fuente de datos original.



**Figura 388.** Transformación STG\_EST\_DOC\_MAT

Esta transformación crea una tabla en la base de datos STG\_DWH, llamada STG\_EST\_DOC\_MAT contiene toda la información del archivo plano CUR\_MAT\_PRO\_PAR.csv fuente de datos original.

Y por ultimo en este proceso se tiene el trabajo que se encargará de ejecutar las transformaciones detalladas anteriormente.



**Figura 399.** Trabajo Job\_STG

Este trabajo ejecuta las transformaciones anteriormente detalladas que realizan la carga ordenada de los datos requeridos en forma secuencial. La ejecución comienza en el nodo Start para luego continuar con el siguiente paso cargar las transformaciones, el trabajo consisten en limpiar los datos de origen cuando es necesario y hacer el correspondiente mapeo en las nuevas tablas. Las transformaciones se componen de pasos y saltos de control o de flujo entre dichos pasos. Estos están clasificados en varios grupos, a modo de ejemplo, citamos: entrada, salida, transformación, flujo, scripting, lookup, validadores, estadísticas, etc.

#### **5.2.2.2 Procesos ETL para carga de datos a las dimensiones**

Este proceso gestiona las transformaciones y trabajos intermedios a ejecutarse y realizan la carga ordenada de los datos requeridos hacia las tablas de dimensiones a utilizarse para la creación de los cubos de información.

A continuación se detallan los procesos creados:

## Tablas Temporales

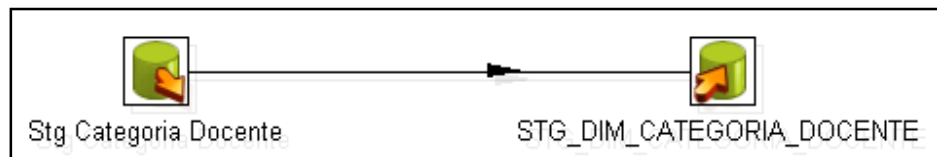


Figura 40. Transformación STG\_DIM\_CATEGORIA\_DOCENTE

Esta transformación crea una tabla en la base de datos STG\_DWH, llamada STG\_DIM\_CATEGORIA\_DOCENTE contiene sólo la información sobre las categorías que tiene el docente cuya fuente de datos es la tabla STG\_CARGA\_HORARIA.



Figura 41. Transformación STG\_DIM\_DEDICACIÓN\_DOCENTE

Esta transformación crea una tabla en la base de datos STG\_DWH, llamada STG\_DIM\_DEDICACION\_DOCENTE contiene sólo la información sobre la dedicación que tiene el docente cuya fuente de datos es la tabla STG\_CARGA\_HORARIA.

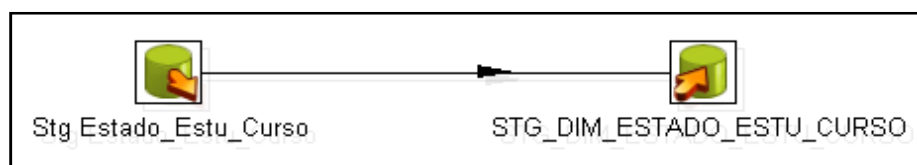
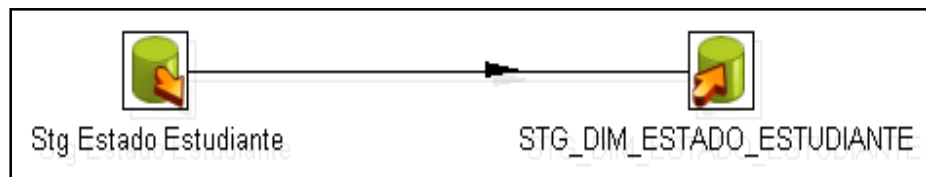


Figura 42. Transformación STG\_DIM\_ESTADO\_ESTU\_CURSO

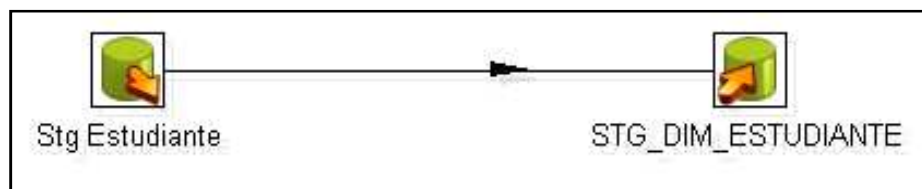
Esta transformación crea una tabla en la base de datos STG\_DWH, llamada STG\_DIM\_ESTADO\_ESTU\_CURSO, contiene sólo la información del estado del estudiante en el periodo lectivos si esta inscrito, matriculado o anulo la matrícula, cuya fuente de datos es la tabla STG\_ESTUDIANTE\_INDICADORES.





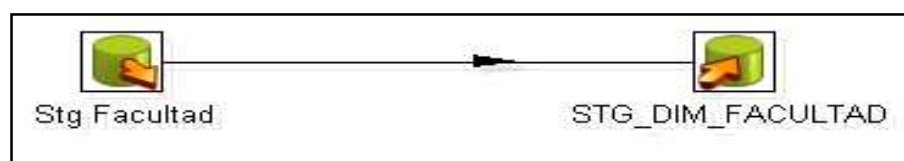
**Figura 43.** Transformación STG\_DIM\_ESTADO\_ESTUDIANTE

Esta transformación crea una tabla en la base de datos STG\_DWH, llamada STG\_DIM\_ESTADO\_ESTUDIANTE, contiene sólo la información del estado en el que se encuentra estudiante si es activo, egresado, cambio de malla, etc., cuya fuente de datos es la tabla STG\_ESTUDIANTE\_INDICADORES.



**Figura 44.** Transformación STG\_DIM\_ESTUDIANTE

Esta transformación crea una tabla en la base de datos STG\_DWH, llamada STG\_DIM\_ESTUDIANTE, contiene toda la información relacionada con el estudiante su nombre, edad, cédula, estado civil, sexo, etc., cuya fuente de datos es la tabla STG\_ESTUDIANTE\_INDICADORES.



**Figura 45.** Transformación STG\_DIM\_FACULTAD

Esta transformación crea una tabla en la base de datos STG\_DWH, llamada STG\_DIM\_FACULTAD, contiene sólo información sobre las carreras, materias de la facultad, cuya fuente de datos son las tablas STG\_ESTUDIANTE\_INDICADORES y STG\_CARGA\_HORARIA.

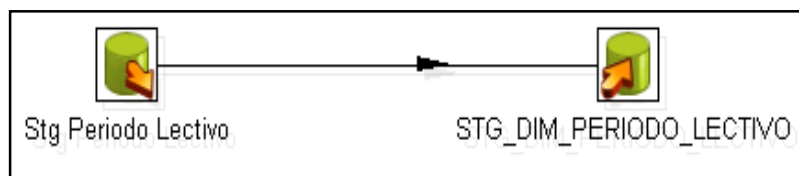


Figura 46. Transformación STG\_DIM\_PERIODO\_LECTIVO

Esta transformación crea una tabla en la base de datos STG\_DWH, llamada STG\_DIM\_PERIODO\_LECTIVO, contiene sólo información sobre los periodos de clases, cuya fuente de datos es la tabla STG\_ESTUDIANTE\_INDICADORES.

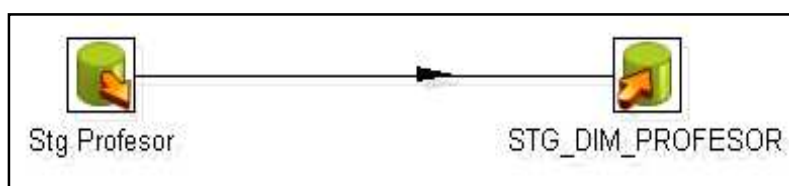


Figura 47. Transformación STG\_DIM\_PROFESOR

Esta transformación crea una tabla en la base de datos STG\_DWH, llamada STG\_DIM\_PROFESOR, contiene información sobre los docentes como nombre, apellido, cédula, cuya fuente de datos es la tabla STG\_CARGA\_HORARIA.

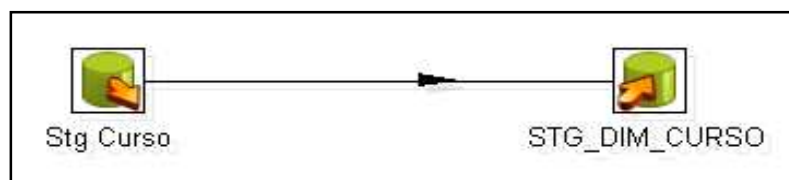


Figura 48. Transformación STG\_DIM\_CURSO

Esta transformación crea una tabla en la base de datos STG\_DWH, llamada STG\_DIM\_CURSO, contiene información sobre los cursos de la facultad, cuya fuente de datos es la tabla STG\_ESTUDIANTE\_INDICADORES.

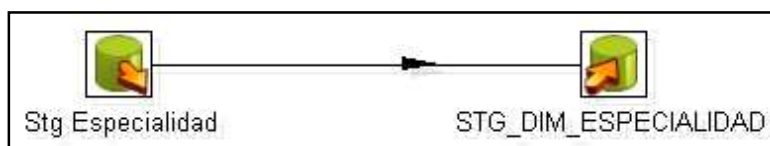


Figura 49. Transformación STG\_DIM\_ESPECIALIDAD

Esta transformación crea una tabla en la base de datos STG\_DWH, llamada STG\_DIM\_ESPECIALIDAD, contiene información sobre las carreras de la facultad, cuya fuente de datos es la tabla STG\_ESTUDIANTE\_INDICADORES.

Y por último en este proceso de las dimensiones se tiene el trabajo que se encargará de ejecutar las transformaciones detalladas anteriormente.

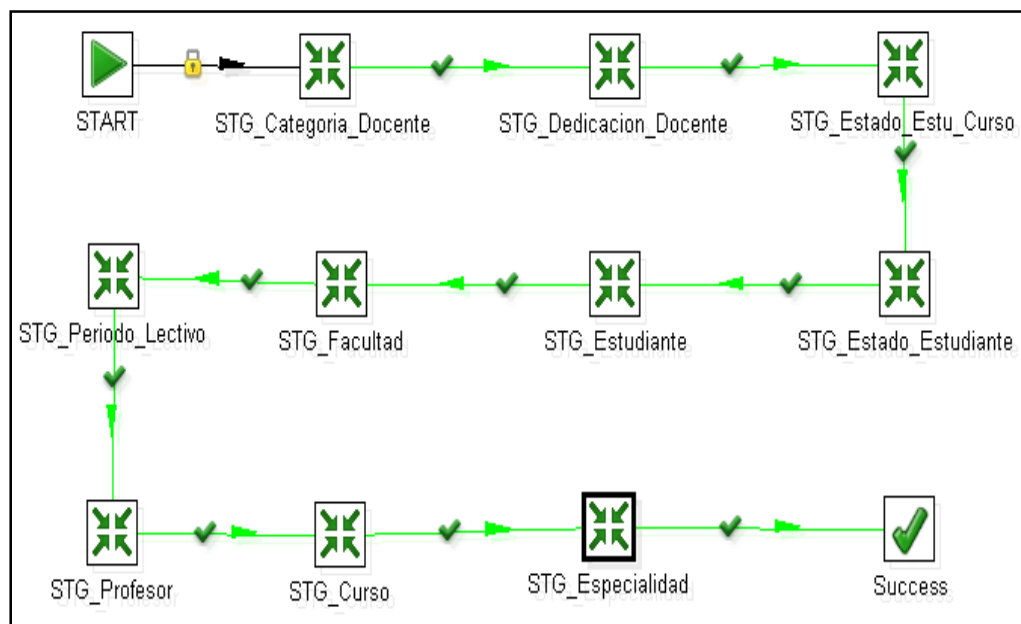


Figura 50. Trabajo Job\_STG\_Dimensiones

Este trabajo ejecuta las transformaciones anteriormente detalladas que realizan la carga ordenada de los datos requeridos en forma secuencial. La ejecución comienza en el nodo Start para luego continuar con el siguiente paso cargar las transformaciones que cargan las tablas temporales que nos permitirán crear las dimensiones.

## Dimensiones

Se empieza el proceso con la carga de la dimensión de fecha ya que esta se corre una sola vez.

Esta transformación crea una tabla en la base de datos DWH, llamada DIM\_FECHA, primero se genera una fecha, luego con el objeto Get Value From

Sequence se genera una secuencia de números empezando por el 1 y luego se sumará a la fecha, luego se calcula la fecha con el objeto Calculator, se continua el proceso filtrando los datos mayores al 31 de diciembre del 2020 con el objeto Filter Rows, se continua llenando la tabla de fecha con otros cálculos adicionales con el objeto Formula, después con el objeto Script Values/Mod se genera un código java script para generar campos adicionales, luego con el objeto Select/Rename values se ordena los campos y por ultimo se carga la tabla DIM\_FECHA.

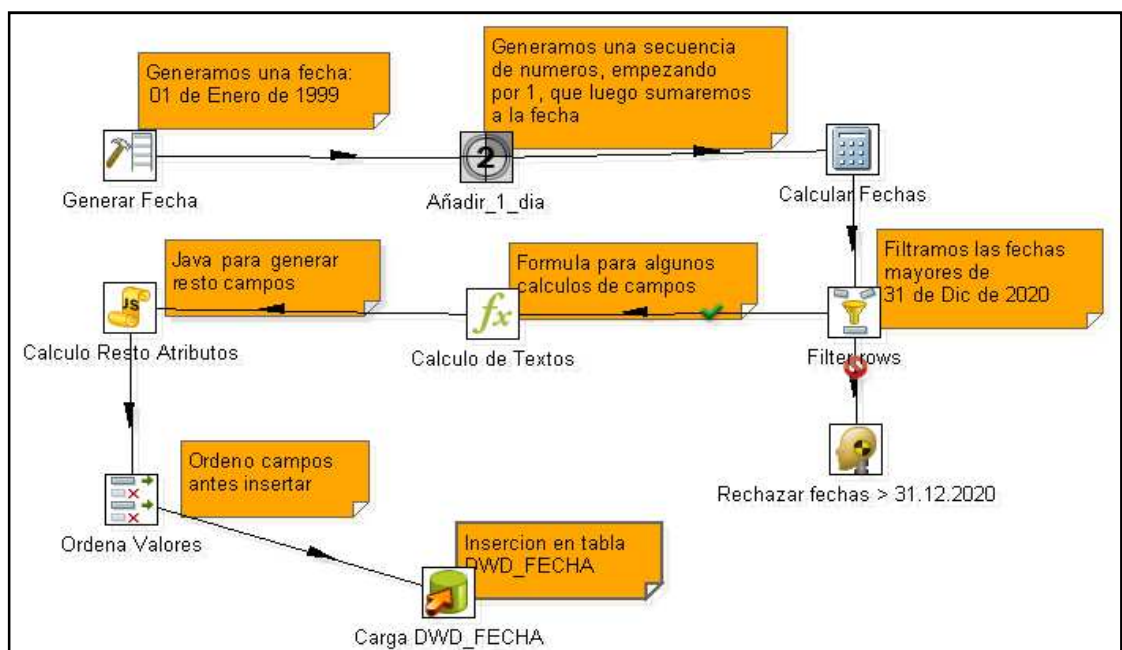


Figura 51. Transformación Dim\_Fecha

Esta transformación no tiene la necesidad de tener un trabajo ya que se ejecuta una sola vez.

A continuación se detallan los procesos creados para las otras dimensiones:

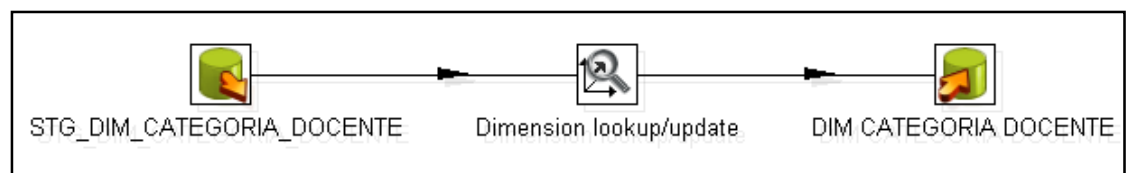


Figura 52. Transformación DIM\_CATEGORIA\_DOCENTE

Esta transformación crea una tabla en la base de datos DWH, llamada DIM\_CATEGORIA\_DOCENTE, se usa el objeto Dimension Lookup para generar las claves subrogadas de la dimensión en base al código de la categoría, la fuente de datos es la tabla STG\_DIM\_CATEGORIA\_DOCENTE.

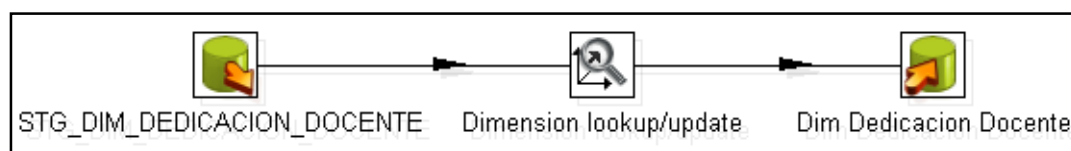


Figura 53. Transformación DIM\_DEDICACION\_DOCENTE

Esta transformación crea una tabla en la base de datos DWH, llamada DIM\_DEDICACION\_DOCENTE, se usa el objeto Dimension Lookup para generar las claves subrogadas de la dimensión en base al código de la dedicación, la fuente de datos es la tabla STG\_DIM\_DEDICACION\_DOCENTE.

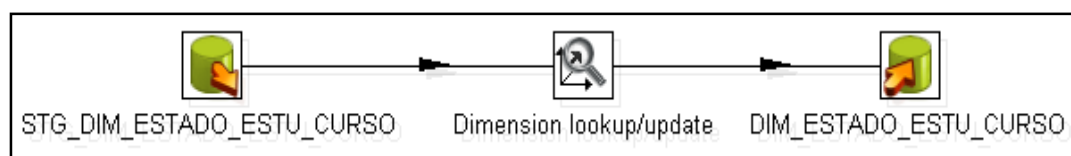


Figura 54. Transformación DIM\_ESTADO\_ESTU\_CURSO

Esta transformación crea una tabla en la base de datos DWH, llamada DIM\_ESTADO\_ESTU\_CURSO, se usa el objeto Dimension Lookup para generar las claves subrogadas de la dimensión en base al código del estado, la fuente de datos es la tabla STG\_DIM\_ESTADO\_ESTU\_CURSO.

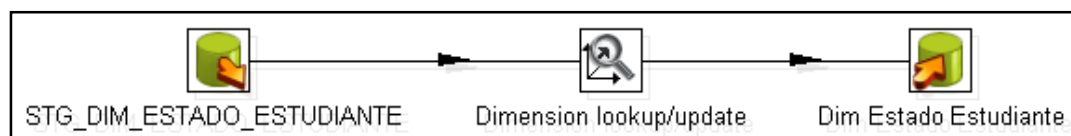


Figura 55. Transformación DIM\_ESTADO\_ESTUDIANTE

Esta transformación crea una tabla en la base de datos DWH, llamada DIM\_ESTADO\_ESTUDIANTE, se usa el objeto Dimension Lookup para generar las claves subrogadas de la dimensión en base al código del estado del estudiante, la fuente de datos es la tabla STG\_DIM\_ESTADO\_ESTUDIANTE.

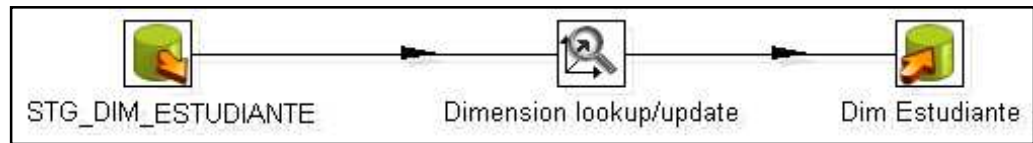


Figura 56. Transformación DIM\_ESTUDIANTE

Esta transformación crea una tabla en la base de datos DWH, llamada DIM\_ESTUDIANTE, se usa el objeto Dimension Lookup para generar las claves subrogadas de la dimensión en base al código del estudiante, la fuente de datos es la tabla STG\_DIM\_ESTUDIANTE.

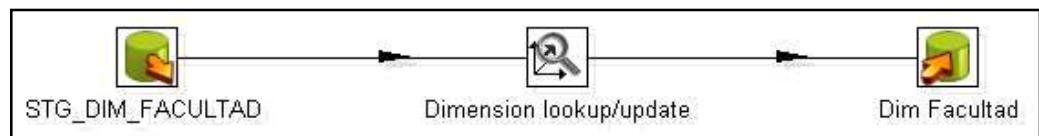


Figura 57. Transformación DIM\_FACULTAD

Esta transformación crea una tabla en la base de datos DWH, llamada DIM\_FACULTAD, se usa el objeto Dimension Lookup para generar las claves subrogadas de la dimensión en base al código de la facultad, escuela y materia, la fuente de datos es la tabla STG\_DIM\_FACULTAD.

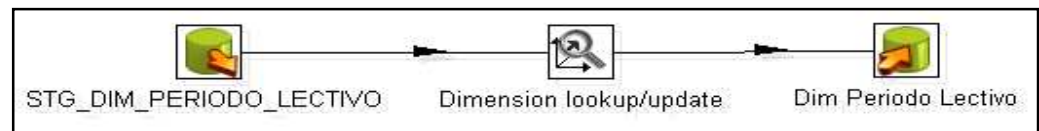


Figura 58. Transformación DIM\_PERIODO\_LECTIVO

Esta transformación crea una tabla en la base de datos DWH, llamada DIM\_PERIODO\_LECTIVO, se usa el objeto Dimension Lookup para generar las claves subrogadas de la dimensión en base al código del periodo lectivo, la fuente de datos es la tabla STG\_DIM\_PERIODO\_LECTIVO.

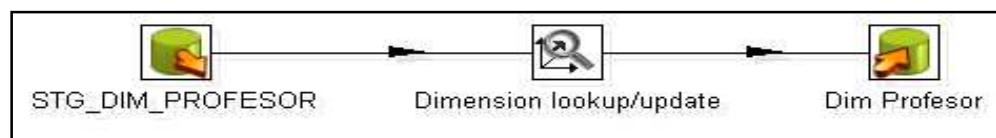


Figura 59. Transformación DIM\_PROFESOR

Esta transformación crea una tabla en la base de datos DWH, llamada DIM\_PROFESOR, se usa el objeto Dimension Lookup para generar las claves subrogadas de la dimensión en base al código del docente, la fuente de datos es la tabla STG\_DIM\_PROFESOR.

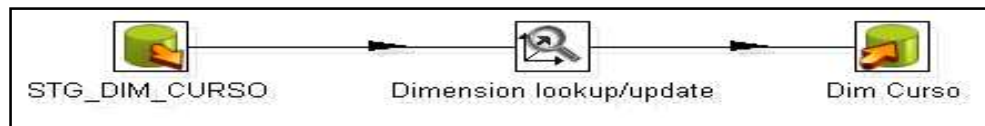


Figura 60. Transformación DIM\_CURSO

Esta transformación crea una tabla en la base de datos DWH, llamada DIM\_CURSO, se usa el objeto Dimension Lookup para generar las claves subrogadas de la dimensión en base al código del curso, la fuente de datos es la tabla STG\_DIM\_CURSO.



Figura 61. Transformación DIM\_ESPECIALIDAD

Esta transformación crea una tabla en la base de datos DWH, llamada DIM\_ESPECIALIDAD, se usa el objeto Dimension Lookup para generar las claves subrogadas de la dimensión en base al código de la facultad, escuela y especialidad, la fuente de datos es la tabla STG\_DIM\_ESPECIALIDAD.

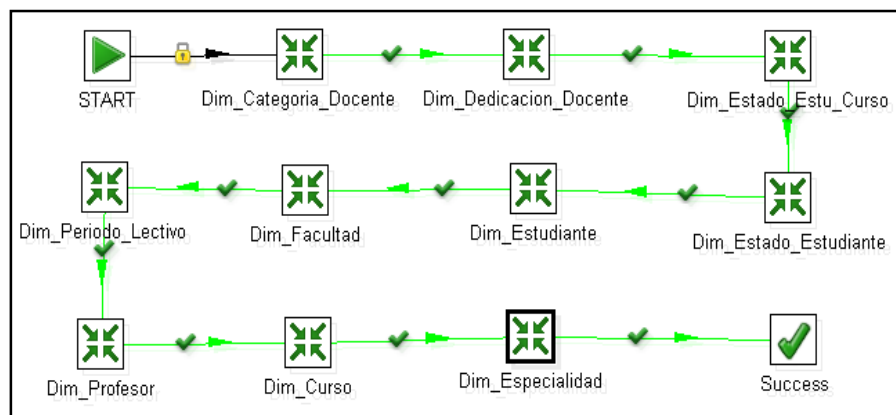


Figura 62. Trabajo Job\_Dimensiones

Este trabajo ejecuta las transformaciones anteriormente detalladas que realizan la carga ordenada de los datos requeridos en forma secuencial. La ejecución comienza en el nodo Start para luego continuar con el siguiente paso cargar las transformaciones que cargan las tablas de las dimensiones que se usarán en la creación de los cubos de información.

### 5.2.2.3 Procesos ETL para carga de datos a los hechos

Este proceso gestiona las transformaciones y trabajos intermedios a ejecutarse y realizan la carga ordenada de los datos requeridos hacia las tablas de hechos a utilizarse para la creación de los cubos de información.

A continuación se detallan los procesos creados:

#### Hecho Estudiante



Figura 63. Transformación TMP\_ESTUDIANTE

Esta transformación crea una tabla en la base de datos DWH, llamada TMP\_ESTUDIANTE, contiene la información relacionada con el estudiante como el número de estudiantes inscritos, la fuente de datos es la tabla STG\_ESTUDIANTE\_INDICADORES.

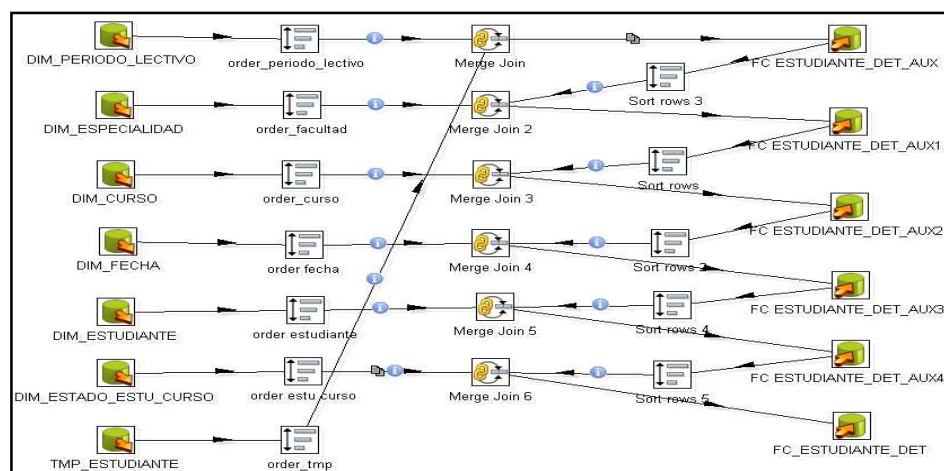


Figura 64. Transformación FC\_ESTUDIANTE\_DETALLE



Esta transformación crea una tabla en la base de datos DWH, llamada FC\_ESTUDIANTE\_DET, la fuente de datos es la tabla TMP\_ESTUDIANTE que se cruza con todas las dimensiones para este modelo, para llegar a la tabla final hay que ordenar las dimensiones y en cascada una a una se van relacionando con la tabla TMP\_ESTUDIANTE creando tablas temporales hasta obtener el hecho final.

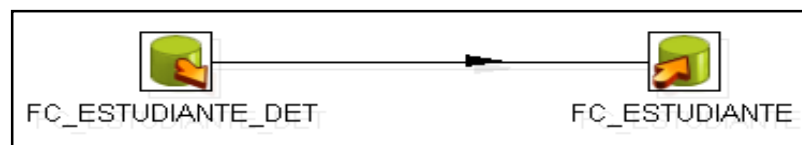


Figura 65. Transformación FC\_ESTUDIANTE

Esta transformación crea una tabla en la base de datos DWH, llamada FC\_ESTUDIANTE, esta tabla tiene sólo los campos de las claves subrogadas y las medidas del hecho FC\_ESTUDIANTE\_DET, la fuente de datos es la tabla FC\_ESTUDIANTE\_DET.

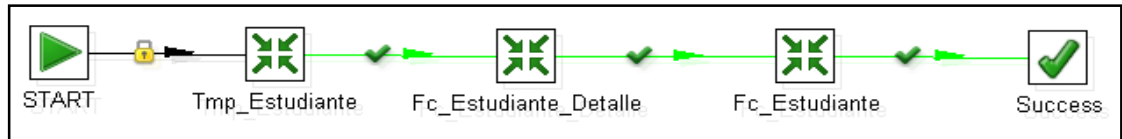


Figura 66. Trabajo Job\_Fc\_Estudiente

Este trabajo ejecuta las transformaciones anteriormente detalladas que realizan la carga ordenada de los datos requeridos en forma secuencial. La ejecución comienza en el nodo Start para luego continuar con el siguiente paso cargar las transformaciones que carga el hecho que se usarán en la creación de los cubos de información.

## Hecho Docente

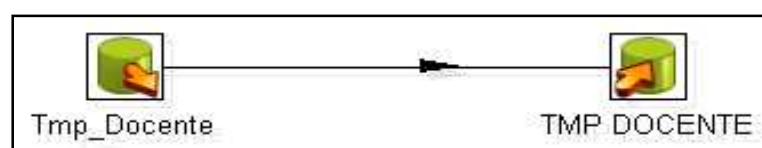


Figura 67. Transformación TMP\_DOCENTE

Esta transformación crea una tabla en la base de datos DWH, llamada TMP\_DOCENTE, contiene la información relacionada con el docente como el número de horas clase, la fuente de datos es la tabla STG\_CARGA\_HORARIA.

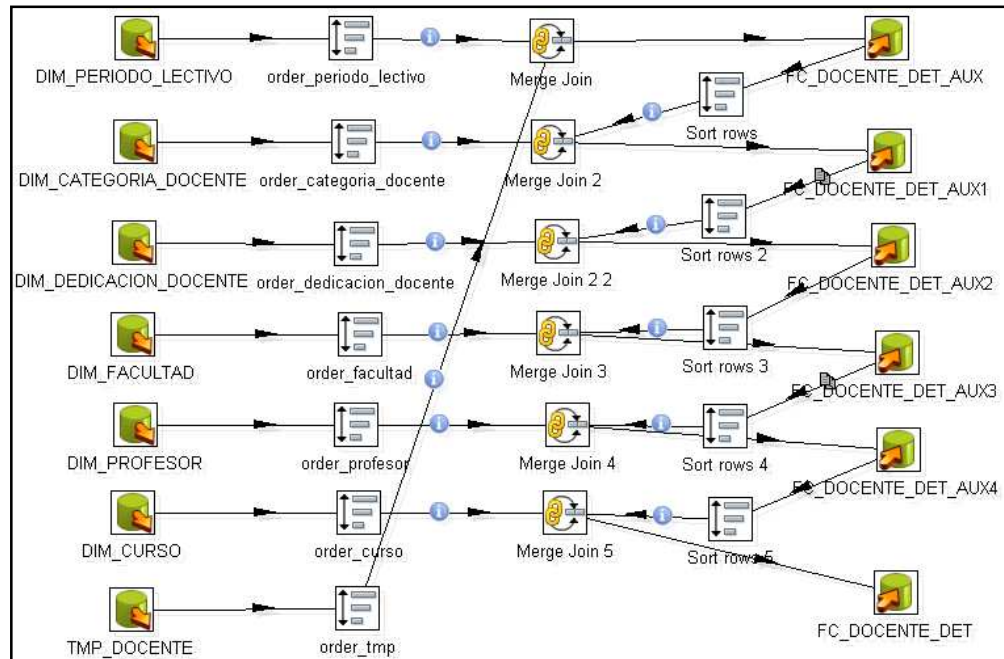


Figura 68. Transformación FC\_DOCENTE\_DETALLE

Esta transformación crea una tabla en la base de datos DWH, llamada FC\_DOCENTE\_DET, la fuente de datos es la tabla TMP\_ DOCENTE que se cruza con todas las dimensiones para este modelo, para llegar a la tabla final hay que ordenar las dimensiones y en cascada una a una se van relacionando con la tabla TMP\_ DOCENTE creando tablas temporales hasta obtener el hecho final.



Figura 69. Transformación FC\_DOCENTE

Esta transformación crea una tabla en la base de datos DWH, llamada FC\_DOCENTE, esta tabla tiene sólo los campos de las claves subrogadas y las medidas del hecho FC\_DOCENTE\_DET, la fuente de datos es la tabla FC\_DOCENTE\_DET.

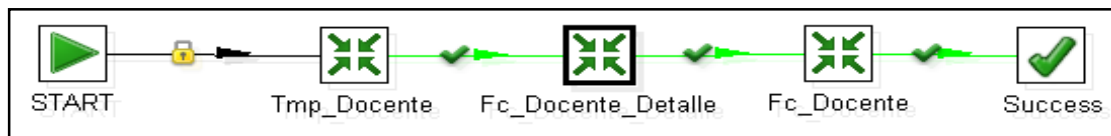


Figura 70. Trabajo Job\_Fc\_Docente

Este trabajo ejecuta las transformaciones anteriormente detalladas que realizan la carga ordenada de los datos requeridos en forma secuencial. La ejecución comienza en el nodo Start para luego continuar con el siguiente paso cargar las transformaciones que carga el hecho que se usarán en la creación de los cubos de información.

### Hecho Docente - Estudiante



Figura 71. Transformación TMP\_DOCENTE-ESTUDIANTE

Esta transformación crea una tabla en la base de datos DWH, llamada TMP\_DOCENTE\_ESTUDIANTE, contiene la información relacionada con el docente y estudiante como el número de estudiantes por docente, la fuente de datos son las tablas TMP\_ESTUDIANTE y TMP\_DOCENTE.

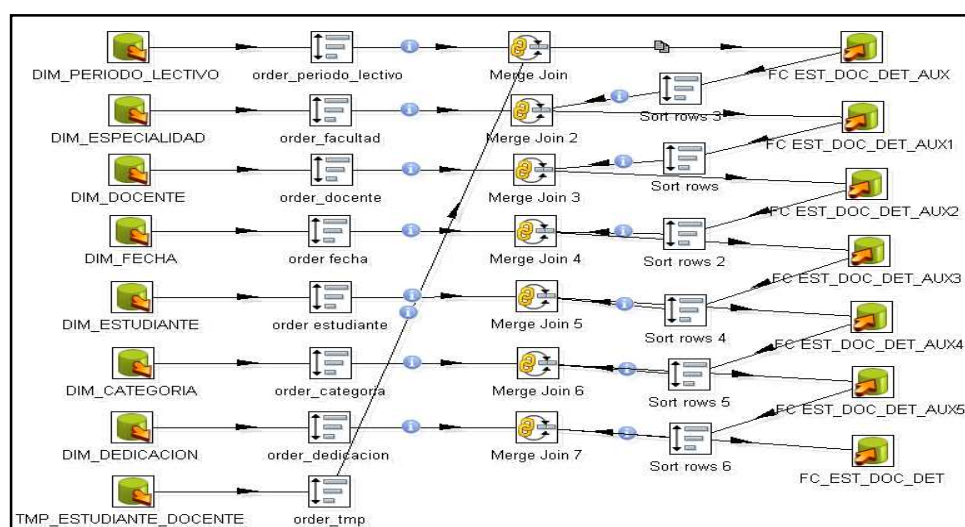


Figura 72. Transformación FC\_ESTUDIANTE-DOCENTE\_DETALLE

Esta transformación crea una tabla en la base de datos DWH, llamada FC\_EST\_DOC\_DET, la fuente de datos es la tabla TMP\_ESTUDIANTE\_DOCENTE que se cruza con todas las dimensiones para este modelo, para llegar a la tabla final hay que ordenar las dimensiones y en cascada una a una se van relacionando con la tabla TMP\_ESTUDIANTE\_DOCENTE creando tablas temporales hasta obtener el hecho final.



Figura 73. Transformación FC\_ESTUDIANTE\_DOCENTE

Esta transformación crea una tabla en la base de datos DWH, llamada FC\_ESTUDIANTE\_DOCENTE, esta tabla tiene sólo los campos de las claves subrogadas y las medidas del hecho FC\_EST\_DOC\_DET, la fuente de datos es la tabla FC\_EST\_DOC\_DET.

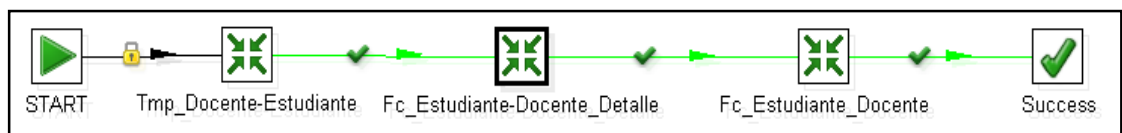


Figura 74. Trabajo Job\_Fc\_Est\_Doc

Este trabajo ejecuta las transformaciones anteriormente detalladas que realizan la carga ordenada de los datos requeridos en forma secuencial. La ejecución comienza en el nodo Start para luego continuar con el siguiente paso cargar las transformaciones que carga el hecho que se usarán en la creación de los cubos de información.

### Hecho Registro Académico

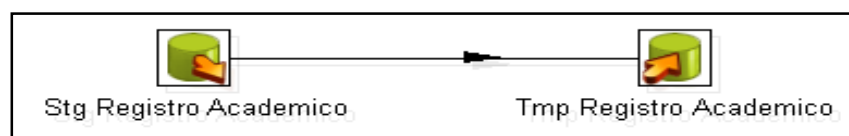


Figura 75. Transformación TMP\_REGISTRO\_ACADEMICO

Esta transformación crea una tabla en la base de datos DWH, llamada TMP\_REGISTRO\_ACADEMICO, contiene la información relacionada con el record académico del estudiante como el número de estudiantes de primera matricula, la fuente de datos es la tabla TMP\_ESTUDIANTE.

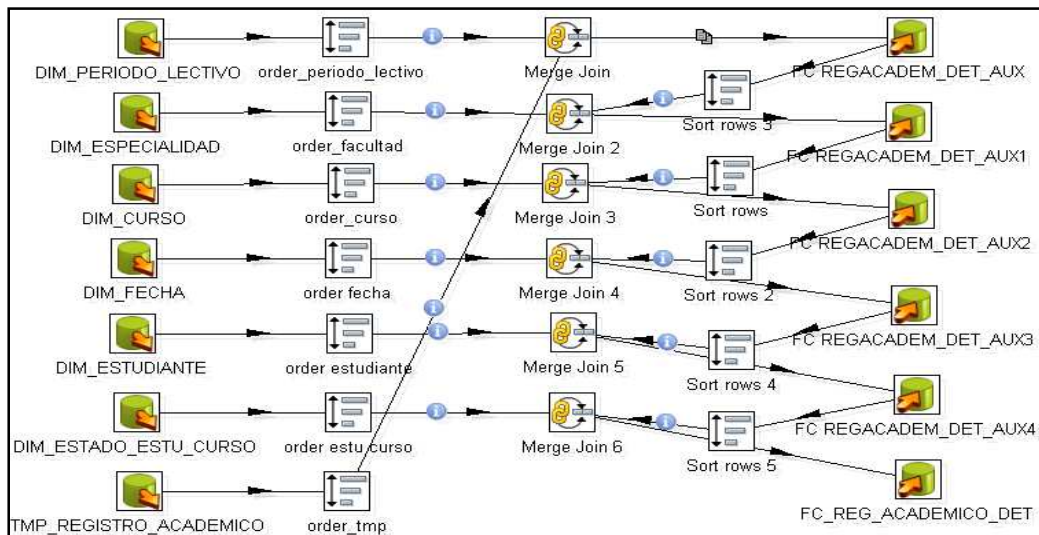


Figura 76. Transformación FC\_REGISTRO\_ACADEMICO\_DETALLE

Esta transformación crea una tabla en la base de datos DWH, llamada FC\_REG\_ACADEMICO\_DET, la fuente de datos es la tabla TMP\_REGISTRO\_ACADEMICO que se cruza con todas las dimensiones para este modelo, para llegar a la tabla final hay que ordenar las dimensiones y en cascada una a una se van relacionando con la tabla TMP\_REGISTRO\_ACADEMICO creando tablas temporales hasta obtener el hecho final.

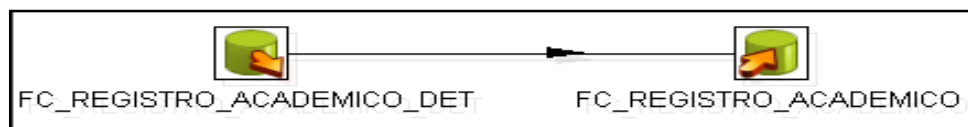


Figura 77. Transformación FC\_REGISTRO\_ACADEMICO

Esta transformación crea una tabla en la base de datos DWH, llamada FC\_REGISTRO\_ACADEMICO, esta tabla tiene sólo los campos de las claves

subrogadas y las medidas del hecho FC\_REG\_ACADEMICO\_DET, la fuente de datos es la tabla FC\_REG\_ACADEMICO\_DET.



Figura 78. Trabajo Job\_Fc\_Registro\_Academico

Este trabajo ejecuta las transformaciones anteriormente detalladas que realizan la carga ordenada de los datos requeridos en forma secuencial. La ejecución comienza en el nodo Start para luego continuar con el siguiente paso cargar las transformaciones que carga el hecho que se usarán en la creación de los cubos de información.

## Hecho Cohorte

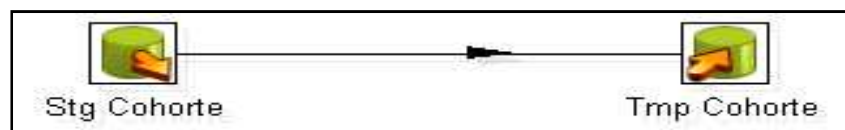


Figura 79. Transformación TMP\_COHORTE

Esta transformación crea una tabla en la base de datos DWH, llamada TMP\_COHORTE, contiene la información relacionada con la cohorte del estudiante como el número de estudiantes de primera matricula, la fuente de datos es la tabla TMP\_REGISTRO\_ACADEMICO.

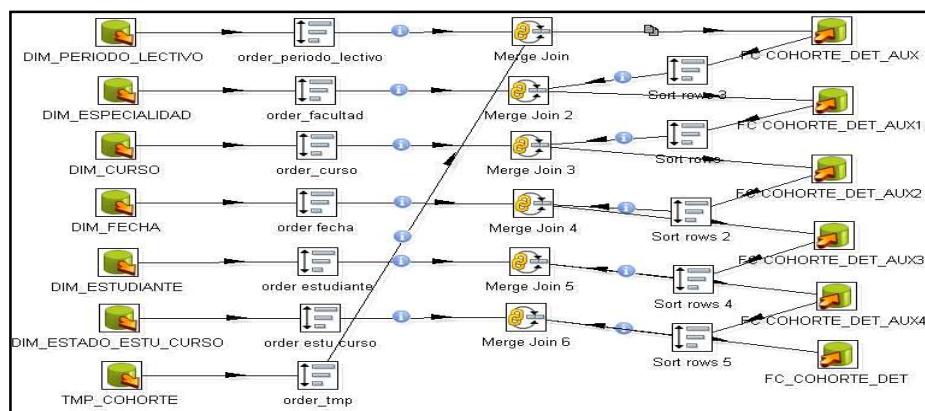


Figura 80. Transformación FC\_COHORTE\_DETALLE

Esta transformación crea una tabla en la base de datos DWH, llamada FC\_COHORTE\_DET, la fuente de datos es la tabla TMP\_COHORTE que se cruza con todas las dimensiones para este modelo, para llegar a la tabla final hay que ordenar las dimensiones y en cascada una a una se van relacionando con la tabla TMP\_COHORTE creando tablas temporales hasta obtener el hecho final.



Figura 81. Transformación FC\_COHORTE

Esta transformación crea una tabla en la base de datos DWH, llamada FC\_COHORTE, esta tabla tiene sólo los campos de las claves subrogadas y las medidas del hecho FC\_COHORTE\_DET, la fuente de datos es la tabla FC\_COHORTE\_DET.



Figura 82. Trabajo Job\_Fc\_Cohorte

Este trabajo ejecuta las transformaciones anteriormente detalladas que realizan la carga ordenada de los datos requeridos en forma secuencial. La ejecución comienza en el nodo Start para luego continuar con el siguiente paso cargar las transformaciones que carga el hecho que se usarán en la creación de los cubos de información.

## Hecho Deserción

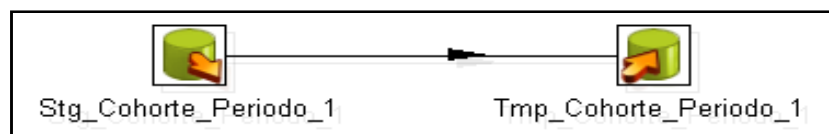


Figura 83. Transformación TMP\_COHORTE\_PERIODO\_1

Esta transformación crea una tabla temporal en la base de datos DWH, llamada TMP\_COHORTE\_PERIODO\_1, contiene la información relacionada con el

estudiante como el número de estudiantes que desertaron de una cohorte determinada, la fuente de datos son las tablas FC\_REG\_ACADEMICO\_DET y FC\_COHORTE\_DET.

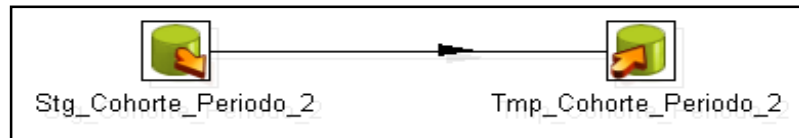


Figura 84. Transformación TMP\_COHORTE\_PERIODO\_2

Esta transformación crea una tabla temporal en la base de datos DWH, llamada TMP\_COHORTE\_PERIODO\_2, contiene la información relacionada con el estudiante como el número de estudiantes que desertaron de una cohorte determinada, la fuente de datos son las tablas FC\_REG\_ACADEMICO\_DET y FC\_COHORTE\_DET.

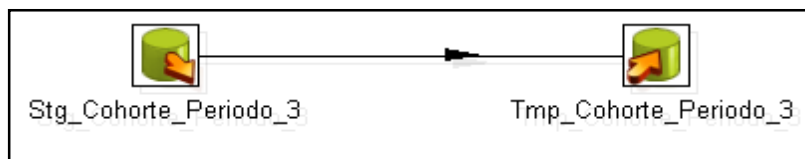


Figura 85. Transformación TMP\_COHORTE\_PERIODO\_3

Esta transformación crea una tabla temporal en la base de datos DWH, llamada TMP\_COHORTE\_PERIODO\_3, contiene la información relacionada con el estudiante como el número de estudiantes que desertaron de una cohorte determinada, la fuente de datos son las tablas FC\_REG\_ACADEMICO\_DET y FC\_COHORTE\_DET.



Figura 86. Trabajo Job\_Tmp\_Periodo

Este trabajo ejecuta las transformaciones de las tablas temporales anteriormente detalladas que realizan la carga ordenada de los datos requeridos en forma secuencial. La ejecución comienza en el nodo Start para luego continuar con el



siguiente paso cargar las transformaciones que carga el hecho que se usarán en la creación de los cubos de información.

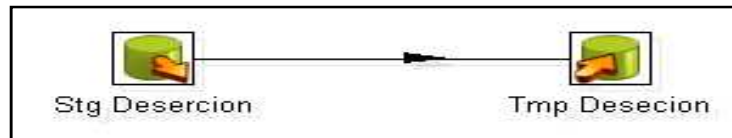


Figura 87. Transformación TMP\_DESERCION

Esta transformación crea una tabla en la base de datos DWH, llamada TMP\_DESERCION, contiene la información relacionada con los estudiante como el número de estudiantes que desertaron en un cohorte, la fuente de datos son las tabla TMP\_COHORTE, TMP\_COHORTE\_PERIODO\_1, TMP\_COHORTE\_PERIODO\_2 y TMP\_COHORTE\_PERIODO\_3.

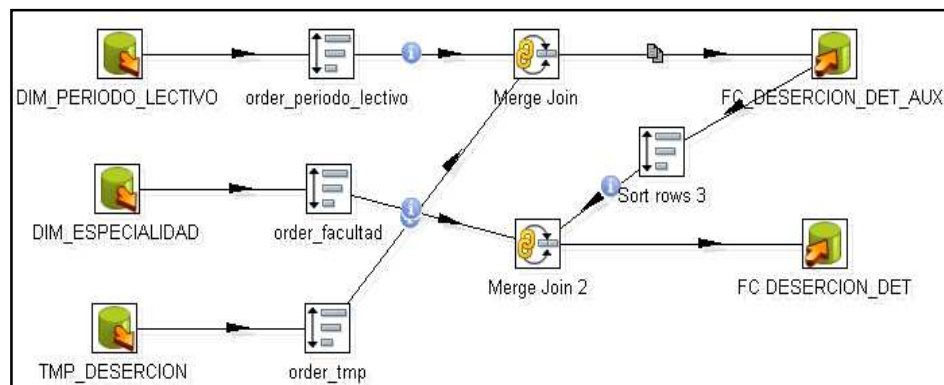


Figura 88. Transformación FC\_DESERCION\_DETALLE

Esta transformación crea una tabla en la base de datos DWH, llamada FC\_DESERCION\_DET, la fuente de datos es la tabla TMP\_DESERCION que se cruza con todas las dimensiones para este modelo, para llegar a la tabla final hay que ordenar las dimensiones y en cascada una a una se van relacionando con la tabla TMP\_DESERCION creando tablas temporales hasta obtener el hecho final.

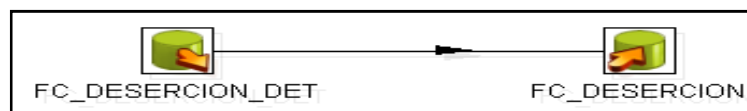


Figura 89. Transformación FC\_DESERCION

Esta transformación crea una tabla en la base de datos DWH, llamada FC\_DESERCION, esta tabla tiene sólo los campos de las claves subrogadas y las medidas del hecho FC\_DESERCION\_DET, la fuente de datos es la tabla FC\_DESERCION\_DET.



Figura 90. Trabajo Job\_Fc\_Desercion

Este trabajo ejecuta las transformaciones anteriormente detalladas que realizan la carga ordenada de los datos requeridos en forma secuencial. La ejecución comienza en el nodo Start para luego continuar con el siguiente paso cargar las transformaciones que carga el hecho que se usarán en la creación de los cubos de información.

#### 5.2.2.4 Procesos de carga y actualización

Las tablas de las dimensiones y hechos que se cargan mediante el proceso de transformaciones y trabajos realizados en Pentaho se ejecutan automáticamente mediante la herramienta Kitchen, este es un programa que ejecuta los trabajos diseñados en Spoon y que se encuentran almacenados como XML ó están en el repositorio de la base de datos.

A continuación se presenta las líneas de código que se crearon para automatizar los trabajos que se usaron en el desarrollo del proyecto:

- Se crearon los siguientes bat para cada trabajo:
  - Ejecuta\_Job\_STG.bat

```
kitchen.bat /rep:pdi_repo /user:admin /pass:admin /dir:/STAGE/jobs  
/job:Job_STG /level:Basic
```
  - Ejecuta\_Job\_STG\_Dimensiones.bat

kitchen.bat /rep:pdi\_repo /user:admin /pass:admin /dir:/STAGE/jobs  
/job:Job\_STG\_Dimensiones /level:Basic

○ Ejecuta\_Job\_Dimensiones.bat

kitchen.bat /rep:pdi\_repo /user:admin /pass:admin  
/dir:/Indicadores/jobs /job:Job\_Dimensiones /level:Basic

○ Ejecuta\_Job\_Fc\_Estudiante.bat

kitchen.bat /rep:pdi\_repo /user:admin /pass:admin  
/dir:/Indicadores/jobs /job:Job\_Fc\_Estudiante /level:Basic

○ Ejecuta\_Job\_Fc\_Docente.bat

kitchen.bat /rep:pdi\_repo /user:admin /pass:admin  
/dir:/Indicadores/jobs /job:Job\_Fc\_Docente /level:Basic

○ Ejecuta\_Job\_Fc\_Est\_Doc.bat

kitchen.bat /rep:pdi\_repo /user:admin /pass:admin  
/dir:/Indicadores/jobs /job:Job\_Fc\_Est\_Doc /level:Basic

○ Ejecuta\_Job\_Fc\_Registro\_Academico.bat

kitchen.bat /rep:pdi\_repo /user:admin /pass:admin  
/dir:/Indicadores/jobs /job:Job\_Fc\_Registro\_Academico /level:Basic

○ Ejecuta\_Job\_Fc\_Cohorte.bat

kitchen.bat /rep:pdi\_repo /user:admin /pass:admin  
/dir:/Indicadores/jobs /job:Job\_Fc\_Cohorte /level:Basic

○ Ejecuta\_Job\_Tmp\_Periodo.bat

kitchen.bat /rep:pdi\_repo /user:admin /pass:admin  
/dir:/Indicadores/jobs /job:Job\_Tmp\_Periodo /level:Basic

○ Ejecuta\_Job\_Fc\_Desercion.bat

```
kitchen.bat      /rep:pdi_repo      /user:admin      /pass:admin  
/dir:/Indicadores/jobs /job:Job_Fc_Desercion /level:Basic
```

- Se creó un bat que genere los bat antes mencionados y se realizó una tarea programada de Windows para que se ejecute el último viernes de cada mes.

- ```
call C:\pentaho\pdi-ce-4.3.0-stable\data-integration\  
Ejecuta_Job_STG.bat  
call C:\pentaho\pdi-ce-4.3.0-stable\data-integration\  
Ejecuta_Job_STG_Dimensiones.bat  
call C:\pentaho\pdi-ce-4.3.0-stable\data-integration\  
Ejecuta_Job_Dimensiones.bat  
call C:\pentaho\pdi-ce-4.3.0-stable\data-integration\  
Ejecuta_Job_Fc_Estudiante.bat  
call C:\pentaho\pdi-ce-4.3.0-stable\data-integration\  
Ejecuta_Job_Fc_Docente.bat  
call C:\pentaho\pdi-ce-4.3.0-stable\data-integration\  
Ejecuta_Job_Fc_Est_Doc.bat  
call C:\pentaho\pdi-ce-4.3.0-stable\data-integration\  
Ejecuta_Job_Fc_Registro_Academico.bat  
call C:\pentaho\pdi-ce-4.3.0-stable\data-integration\  
Ejecuta_Job_Fc_Cohorte.bat  
call C:\pentaho\pdi-ce-4.3.0-stable\data-integration\  
Ejecuta_Job_Tmp_Periodo.bat  
call C:\pentaho\pdi-ce-4.3.0-stable\data-integration\  
Ejecuta_Job_Fc_Desercion.bat
```

### 5.3 Cubos y Análisis OLAP

La creación de los cubos de información se basa en tres pasos, el primer paso es decidir cuáles son los atributos que son requeridos para realizar el análisis, así como también el diseño del tipo de estructura lógica que se va a requerir.

Como segundo paso se debe especificar la estructura de las dimensiones diseñando las agregaciones junto con la creación de todas las tablas relacionales requeridas, y como último se debe realizar la carga del cubo para procesarlo.

Para poder realizar análisis OLAP en Pentaho se deben tener primero los cubos de información de Mondrian. Los cubos que utiliza este servidor son llamados esquemas y son archivos XML que definen las dimensiones que va a tener el cubo. Dichos esquemas pueden ser generados de forma manual o a través de la herramienta de Schema Workbench de la plataforma. En este proyecto se utilizó la herramienta de Pentaho antes mencionada para la creación de los cubos.

### **5.3.1 Estructura General**

En Pentaho los cubos se especifican mediante un archivo donde se describe la estructura de los mismos (dimensiones, jerarquías, niveles, hechos, medidas) así como se realizan los mapeos entre estas estructuras y las tablas donde se encuentran almacenados los datos que pueblan las mismas.

A continuación se presenta la estructura general de los cubos utilizados dentro del proyecto. Como se pueden ver en las figuras se denota que en la herramienta la estructura del cubo se detalla como una estructura de árbol en la cual a partir del nodo padre Cubo se desprenden cada uno de los componentes que lo definen.

#### **Cubo Docentes**

El cubo de docentes de nombre Docentes que se ve en la figura contiene un conjunto de dimensiones que se encuentran en el primer nivel, también la tabla de hechos de la misma y por último las medidas calculadas que sean necesarias para la obtención de los indicadores académicos.

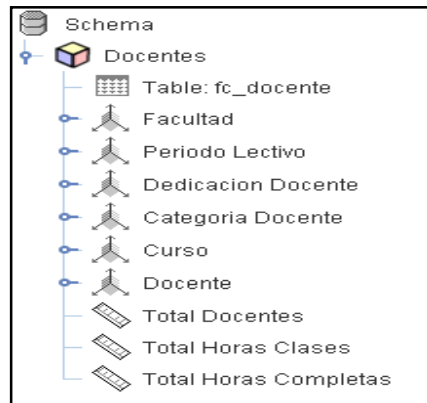


Figura 91. Representación Cubo Docentes

Desplegando la estructura correspondiente al diseño de la dimensión Facultad, se puede observar la representación de una jerarquía interna de cuatro niveles. Estos son Facultad, Escuela, Carrera y Materia, el resto de dimensiones solamente tienen un nivel.

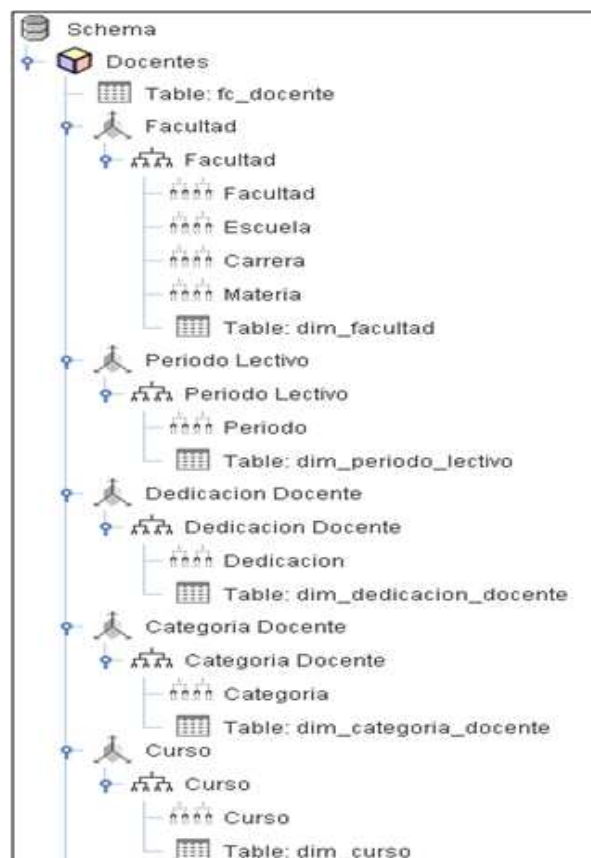




Figura 92. Representación Cubo Docentes Detallado

### Cubo Estudiantes

El cubo de estudiantes de nombre Estudiantes que se ve en la figura contiene un conjunto de dimensiones que se encuentran en el primer nivel, también la tabla de hechos de la misma y por último la medida calculada que es necesaria para la obtención de los indicadores académicos.

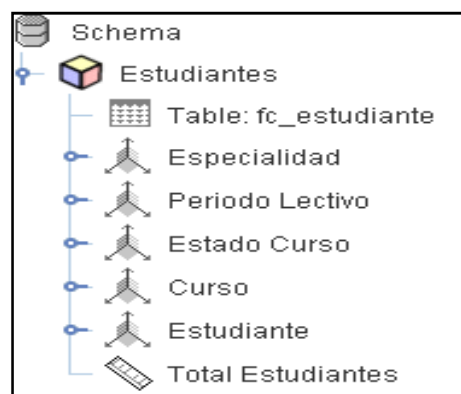


Figura 93. Representación Cubo Estudiantes

Desplegando la estructura correspondiente al diseño de la dimensión Especialidad, se puede observar la representación de una jerarquía interna de tres niveles. Estos son Facultad, Escuela y Carrera, el resto de dimensiones solamente tienen un nivel.

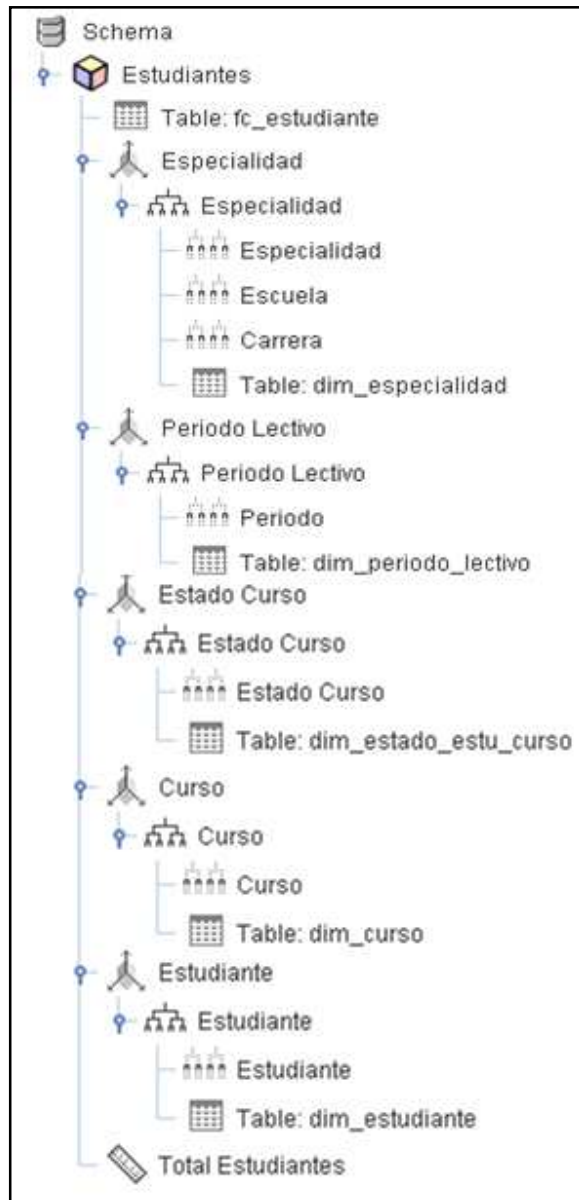


Figura 94. Representación Cubo Estudiante Detallado

### Cubo Docente - Estudiante

El cubo de docentes y estudiantes de nombre Docente-Estudiante que se ve en la figura contiene un conjunto de dimensiones que se encuentran en el primer nivel, también la tabla de hechos de la misma y por último las medidas calculadas que son necesarias para la obtención de los indicadores académicos.



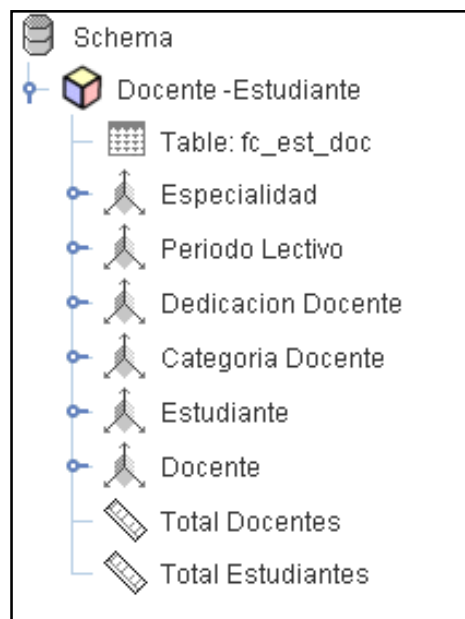
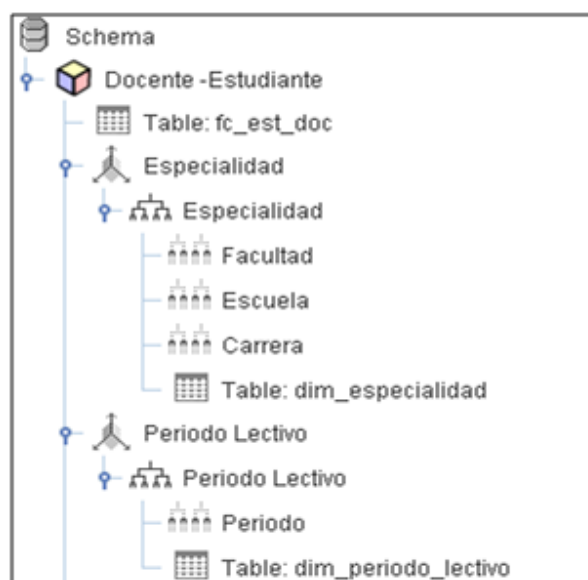


Figura 95. Representación Cubo Docente-Estudiante

Desplegando la estructura correspondiente al diseño de la dimensión Especialidad, se puede observar la representación de una jerarquía interna de tres niveles. Estos son Facultad, Escuela y Carrera, el resto de dimensiones solamente tienen un nivel.



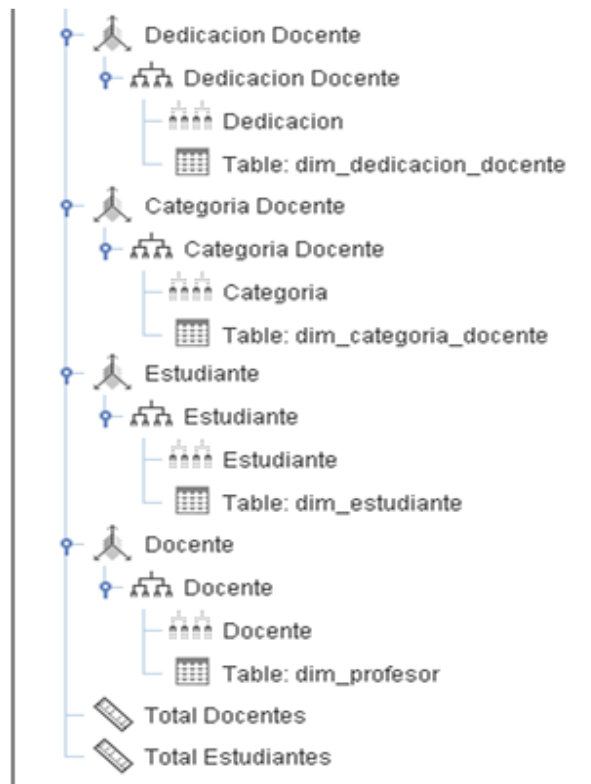


Figura 96. Representación Cubo Docente-Estudiante Detallado

### Cubo Registro Académico

El cubo del registro académico de nombre Registro Académico que se ve en la figura contiene un conjunto de dimensiones que se encuentran en el primer nivel, también la tabla de hechos de la misma y por último las medidas calculadas que son necesarias para la obtención de los indicadores académicos.

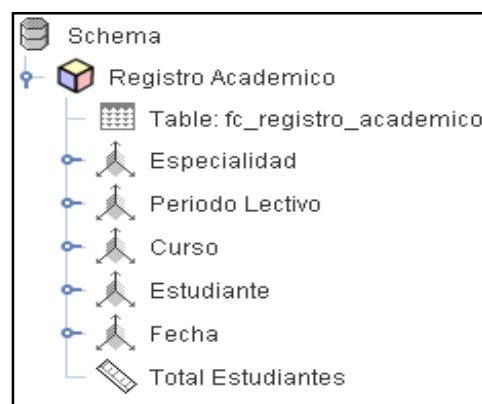


Figura 97. Representación Cubo Registro Académico

Desplegando la estructura correspondiente al diseño de la dimensión Especialidad, se puede observar la representación de una jerarquía interna de tres niveles. Estos son Facultad, Escuela y Carrera; la dimensión de Fecha también se puede observar otra jerarquía de tres niveles. Estos son Año, Semestre, Mes; el resto de dimensiones solamente tienen un nivel.

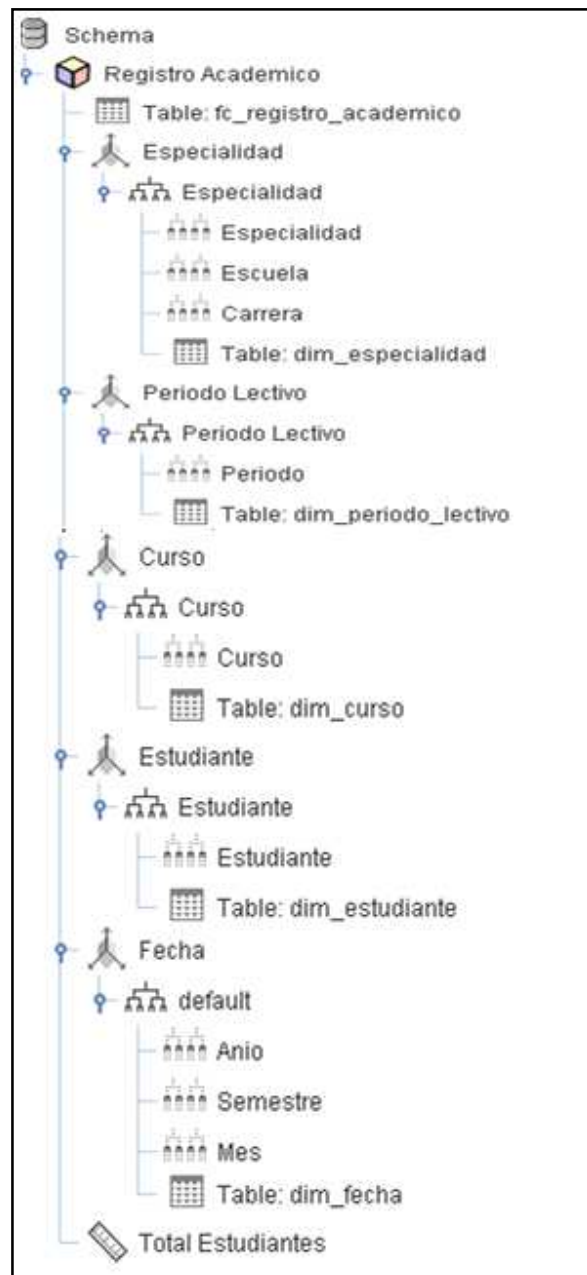


Figura 98. Representación Cubo Registro Académico Detallado

## Cubo Cohorte

El cubo de la cohorte de nombre Cohorte que se ve en la figura contiene un conjunto de dimensiones que se encuentran en el primer nivel, también la tabla de hechos de la misma y por último las medidas calculadas que son necesarias para la obtención de los indicadores académicos.

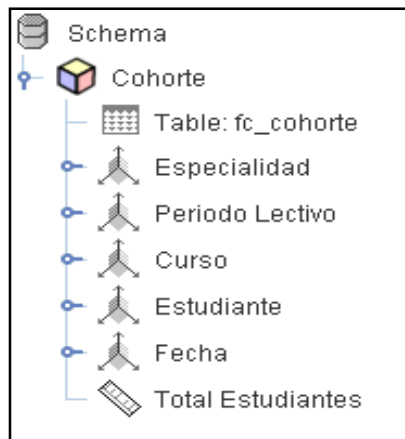
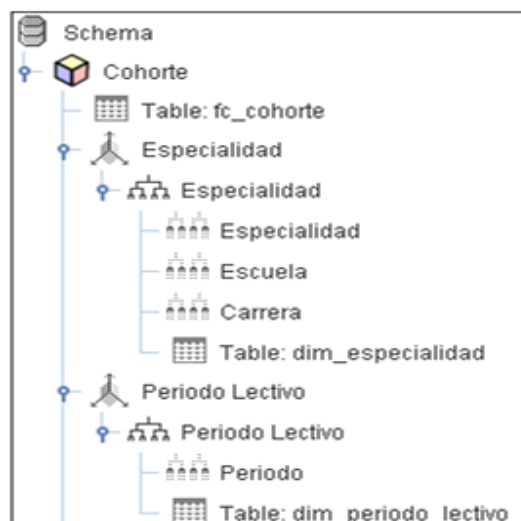


Figura 99. Representación Cubo Cohorte

Desplegando la estructura correspondiente al diseño de la dimensión Especialidad, se puede observar la representación de una jerarquía interna de tres niveles. Estos son Facultad, Escuela y Carrera; en la dimensión de Fecha también se puede observar otra jerarquía de tres niveles. Estos son Año, Semestre, Mes; el resto de dimensiones solamente tienen un nivel.



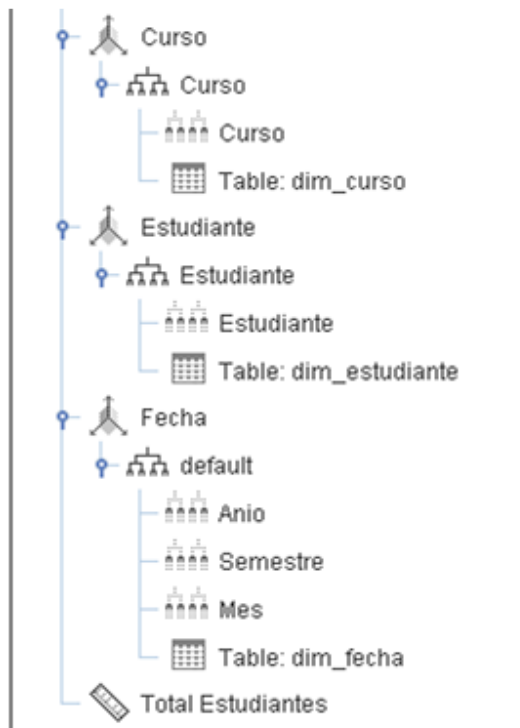


Figura 100. Representación Cubo Cohorte Detallado

## Cubo Deserción

El cubo de la deserción de nombre Deserción que se ve en la figura contiene un conjunto de dimensiones que se encuentran en el primer nivel, también la tabla de hechos de la misma y por último las medidas calculadas que son necesarias para la obtención de los indicadores académicos.

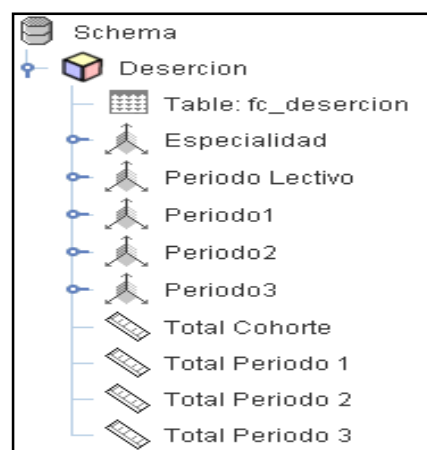


Figura 101. Representación Cubo Deserción

Desplegando la estructura correspondiente al diseño de la dimensión Especialidad, se puede observar la representación de una jerarquía interna de tres niveles. Estos son Facultad, Escuela y Carrera; el resto de dimensiones solamente tienen un nivel.

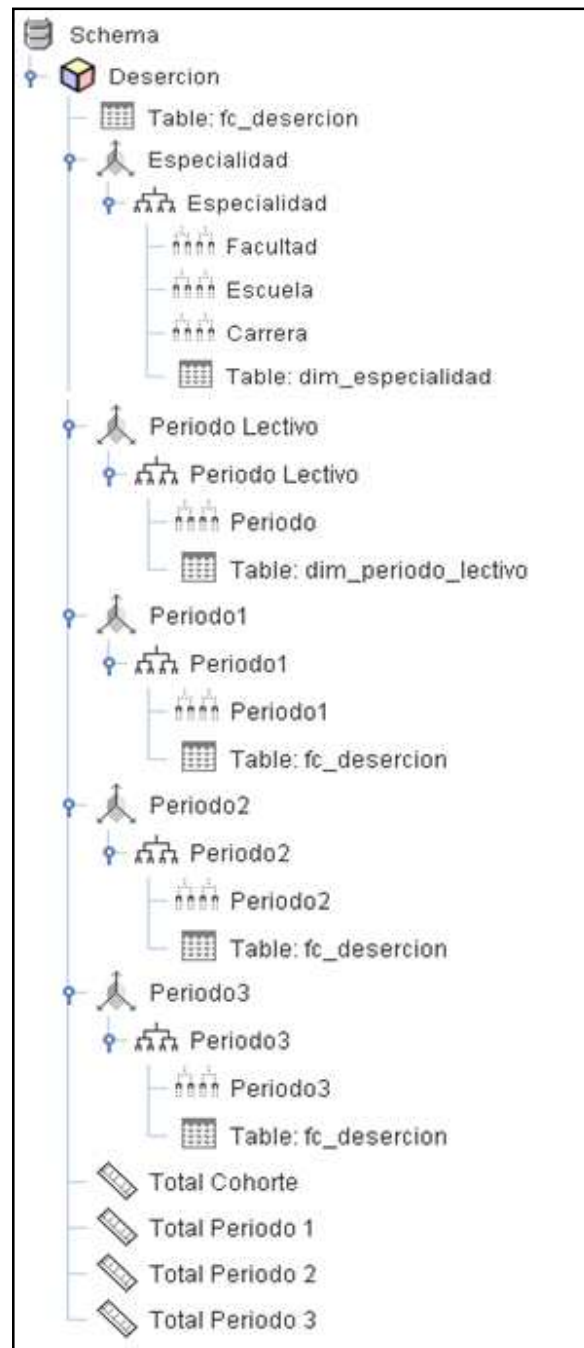


Figura 102. Representación Cubo Deserción Detallado

### **5.3.2 Publicación Cubos**

Para que los cubos sean accedidos por el motor de cubos estos deben publicarse dentro de la plataforma Pentaho. Existen dos maneras de publicar los cubos en la aplicación, la primera es copiando los archivos XML de definición en el repositorio de esquemas del servidor de cubos Mondrian dentro de la aplicación y luego modificando los archivos de configuración que contienen los datasources del mismo.

Este archivo de configuración indica al motor de cubos donde se encuentra la metadata de los mismos, la cual le informa cuales son las tablas de la base de datos a utilizar al momento de la ejecución.

Esto debe realizarse cada vez que se crea un cubo la cual lo hace bastante engorroso para el usuario final. La segunda manera y como se publicaron los cubos en este proyecto es utilizando la funcionalidad de la herramienta, lo cual es bastante más directo y sencillo.

### **5.4 Reportes e Indicadores**

El problema a resolver dentro de este proyecto fue la generación de reportes e indicadores académicos creados a partir de los datos registrados, de manera que los reportes e indicadores sean usados para la toma de decisiones de la facultad.

La herramienta usada para la creación de los indicadores y reportes es Pentaho Report Designer este es el componente de Pentaho utilizado para la creación de los reportes.

Para la creación de cada reporte se siguieron los siguientes pasos:

- Se configuró la fuente de datos en la cual se agregan las conexiones y demás configuraciones.
- En el Pentaho Analysis Schema File se asigna el archivo XML conteniendo el diseño del cubo a utilizar.

- Se asigna la conexión a la base de datos donde se encuentra cargado el Data Mart.
- De la vista del análisis se obtiene la consulta MDX utilizada para obtener los datos a presentar.
- Se seleccionan las columnas a mostrar.
- Luego se pasa a la etapa de personalización del reporte.
- El último paso es la publicación del reporte dentro del servidor para que pueda ser visualizado.

#### **5.4.1 Estructura de los Reportes e Indicadores**

Una definición general de un reporte consiste en un conjunto de secciones que definen la disposición y contenido de la información dentro de éste. Estas secciones son:

- Cabecera y pie del reporte: Es impreso al comienzo y fin del reporte respectivamente.
- Cabecera y pie de página: Son impresos al comienzo y fin de cada página respectivamente.
- Cabecera y pie de grupo: Son impresos al comienzo y fin de cada grupo respectivamente. Un grupo, generalmente, contiene el nombre de una columna y su valor.
- Ítems o detalles: Contienen los datos obtenidos de la consulta. Estos valores se repiten tantas veces como las devuelva en la consulta.
- Sección de funciones y expresiones: Permiten realizar cálculos de valores. Por ejemplo se podría calcular el total de un valor que pertenece a un grupo.

A continuación se detalla la estructura general de los reportes e indicadores:



## Reportes Horario Docente

Se crearon tres reportes con la misma estructura para cada una de las escuelas que tiene la facultad ciencias, civil y geomensura.

El reporte esta compuesto por tres partes:

- La cabecera que contiene el nombre de la facultad y la escuela.
- El agrupamiento de la cabecera que tiene el nombre del docente su dedicación y categoría.
- El cuerpo genera información de la materia que imparte el docente, el día, hora, aula, curso, paralelo y la carga horaria que tiene el docente por materia.
- Se puede filtrar el reporte por carrera, nombre del docente y se puede seleccionar el formato de salida que desee tenga el reporte.

▼ Report Parameters

Carrera

INFORMATICA

Docente

GOYES MOSQUERA JAIRO JAVIER

Output Type

HTML (Single Page)

View Report

☒ Auto-Submit



**Facultad de Ingeniería Ciencias, Físicas y Matemática**  
**ESCUELA DE CIENCIAS**  
**PROGRAMA ACADEMICO DOCENTE**

| DOCENTE                     | DEDICACION     | CATEGORIA           |      |                    |                  |
|-----------------------------|----------------|---------------------|------|--------------------|------------------|
| GOYES MOSQUERA JAIRO JAVIER | Tiempo Parcial | Auxiliar            |      |                    |                  |
| MATERIA                     | DIA            | HORA                | AULA | CURSO/<br>PARALELO | CARGA<br>HORARIA |
| ALGORITMOS                  | Lunes          | 19:00:00 - 21:00:00 | R-8  | 3 - 1              | 2                |
| ALGORITMOS                  | Martes         | 17:00:00 - 19:00:00 | R-11 | 3 - 1              | 2                |
| ALGORITMOS                  | Jueves         | 17:00:00 - 19:00:00 | S-3  | 3 - 1              | 2                |

Figura 103. Reporte Horario Docente

## Reportes Estudiante por Docente

El reporte está compuesto por dos partes:

- La cabecera que contiene el nombre de la facultad y la escuela.
- El cuerpo genera información de la materia que imparte el docente, curso, paralelo, número de estudiantes en el sistema Fing., número de estudiantes en el sistema SAU y el total de estudiantes.
- Se puede filtrar el reporte por carrera y se puede seleccionar el formato de salida que desee tenga el reporte.


| Report Parameters                                                                   |          |                     |                         |           |          |            |
|-------------------------------------------------------------------------------------|----------|---------------------|-------------------------|-----------|----------|------------|
| Carrera                                                                             |          |                     |                         |           |          |            |
| Ing. Informática                                                                    |          |                     |                         |           |          |            |
| Output Type                                                                         |          |                     |                         |           |          |            |
| HTML (Paginated)                                                                    |          |                     |                         |           |          |            |
| View Report Auto-Submit                                                             |          |                     |                         |           |          |            |
| Page 1 of 7                                                                         |          |                     |                         |           |          |            |
|  |          |                     |                         |           |          |            |
| Facultad de Ingeniería Ciencias, Físicas y Matemática                               |          |                     |                         |           |          |            |
| ESCUELA DE CIENCIAS                                                                 |          |                     |                         |           |          |            |
| REPORTE NUMERO ESTUDIANTES POR DOCENTE                                              |          |                     |                         |           |          |            |
| Carrera: Ing. Informática                                                           |          |                     |                         |           |          |            |
| CURSO                                                                               | PARALELO | MATERIA             | DOCENTE                 | EST. FING | EST. SAU | TOTAL EST. |
| 1                                                                                   | 1        | FÍSICA APLICADA     | ING. FLORES SUAREZ ..   | 0         | 43       | 43         |
| 1                                                                                   | 1        | PROGRAMACION I      | ING. FLORES SUAREZ ..   | 0         | 30       | 30         |
| 1                                                                                   | 1        | ECOLOGÍA EN LAS TIC | LCDA. AVILA SALEM ..    | 1         | 29       | 30         |
| 1                                                                                   | 1        | ANÁLISIS I          | LCDO. FRANCO GUERRER .. | 1         | 38       | 39         |
| 1                                                                                   | 1        | COMUNICACION ORA..  | LCDO. GUERRERO BARRO .. | 0         | 17       | 17         |
| 1                                                                                   | 1        | FUNDAMENTOS DE ..   | MAT. GALINDO DE LA ..   | 0         | 28       | 28         |

Figura 104. Reporte Estudiante por Docente

## Indicadores Carga Horaria Cuerpo Docente

Se crearon tres reportes con la misma estructura para cada una de las escuelas que tiene la facultad ciencias, civil y geomensura.

El reporte está compuesto por tres partes:

- La cabecera que contiene el nombre de la facultad y la escuela.
- El agrupamiento de la cabecera que tiene el nombre de la carrera y periodo a consultar.
- El cuerpo genera información de los indicadores Carga horaria docente y Carga horaria docente / docente, su puede visualizar la fórmula que se utilizó para el cálculo y un gráfico de barras que muestra la variación del indicador en base al periodo.
- Se puede filtrar el reporte por carrera, periodo y se puede seleccionar el formato de salida que desee tenga el reporte.

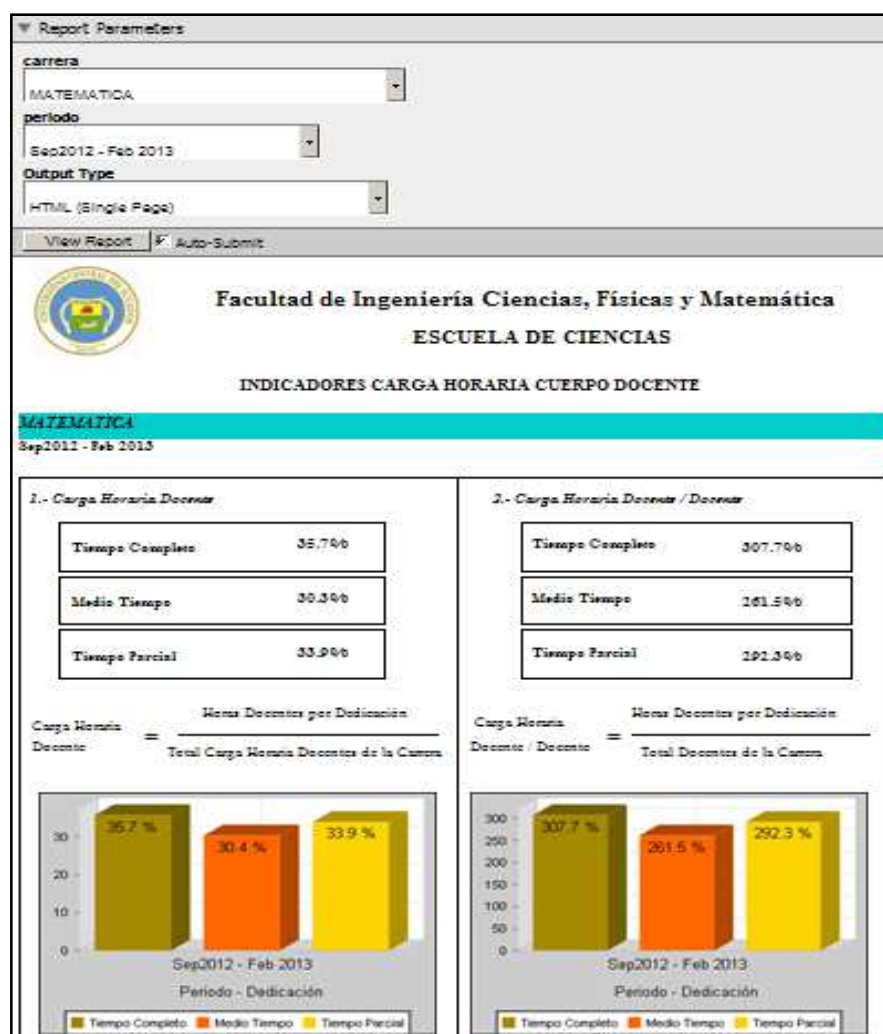


Figura 105. Indicadores Carga Horaria Cuerpo Docente

## **Indicadores Docente**

Se crearon tres reportes con la misma estructura para cada una de las escuelas que tiene la facultad ciencias, civil y geomensura.

El reporte esta compuesto por tres partes:

- La cabecera que contiene el nombre de la facultad y la escuela.
- El agrupamiento de la cabecera que tiene el nombre de la carrera y periodo a consultar.
- El cuerpo genera información de los indicadores Docentes y Docentes / Estudiantes, su puede visualizar la fórmula que se utilizó para el cálculo y un gráfico de barras que muestra la variación del indicador en base al periodo.
- Se puede filtrar el reporte por carrera, periodo y se puede seleccionar el formato de salida que desee tenga el reporte.

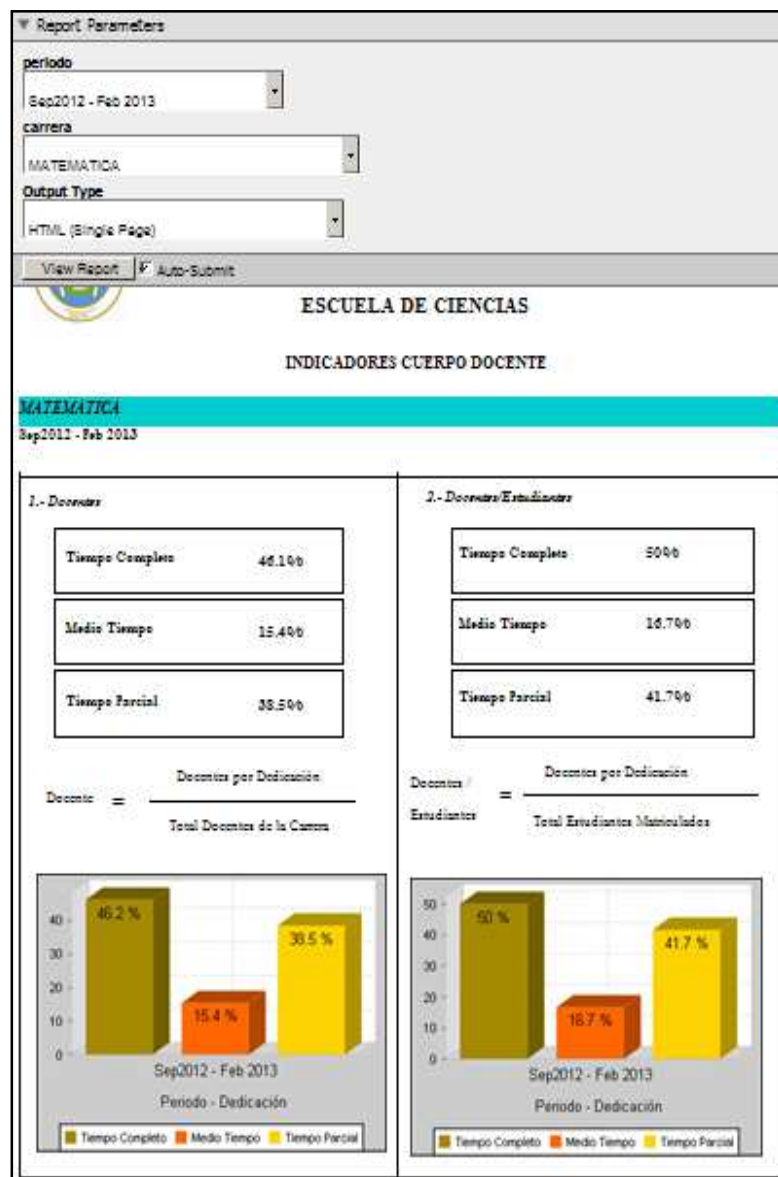


Figura 106. Indicadores Docente

### Indicador Tasa de Deserción

Se crearon tres reportes con la misma estructura para cada una de las escuelas que tiene la facultad ciencias, civil y geomensura.

El reporte esta compuesto por tres partes:

- La cabecera que contiene el nombre de la facultad y la escuela.
- El cuerpo genera información de los estudiantes que desertaron por carrera muestra su cohorte y los tres periodos siguientes y la tasa de deserción, su puede visualizar la fórmula que se utilizó para el cálculo y un gráfico de barras que muestra la variación del indicador en base a la carrera.
- Se puede filtrar el reporte por periodo y se puede seleccionar el formato de salida que desee tenga el reporte.

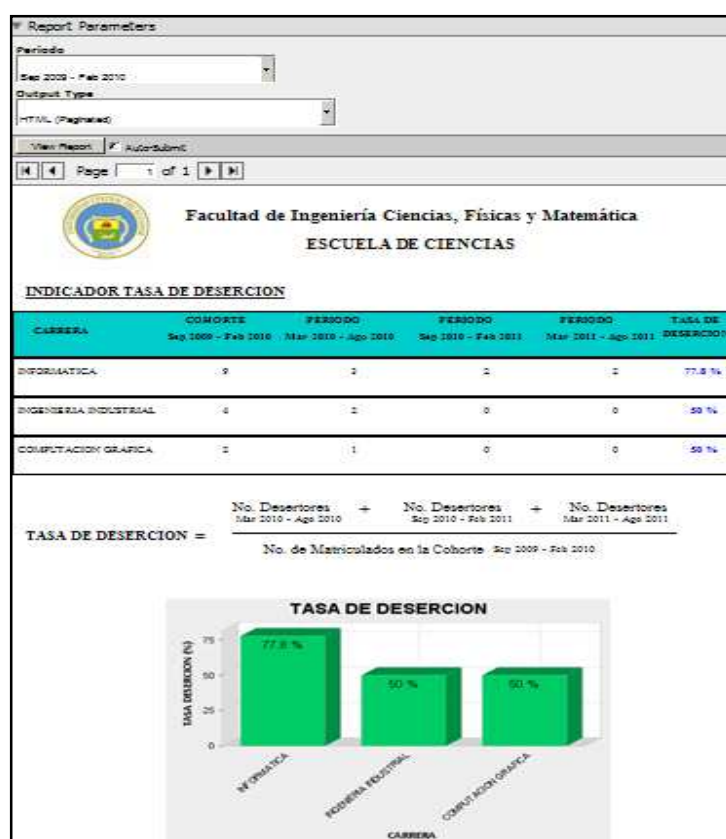


Figura 107. Indicadores Docente

## Indicador Tasa de Deserción

Se crearon tres reportes con la misma estructura para cada una de las escuelas que tiene la facultad ciencias, civil y geomensura.

## 5.5 Administración de Usuarios y Permisos

En la consola de administración es donde se gestionan los usuarios y los roles de cada uno de ellos. Antes de la configuración de los roles se debe planificar como se van a realizar los accesos a los objetos, se debe determinar que roles van a tener sentido dentro de la aplicación, se debe determinar cuales roles van a tener acceso a las URL particulares y determinar si algunos de ellos van a tener capacidad de ejecutar secuencias de acción en el repositorio de soluciones.

La siguiente figura muestra en la consola de administración la pestaña de gestión de usuarios en la aplicación.

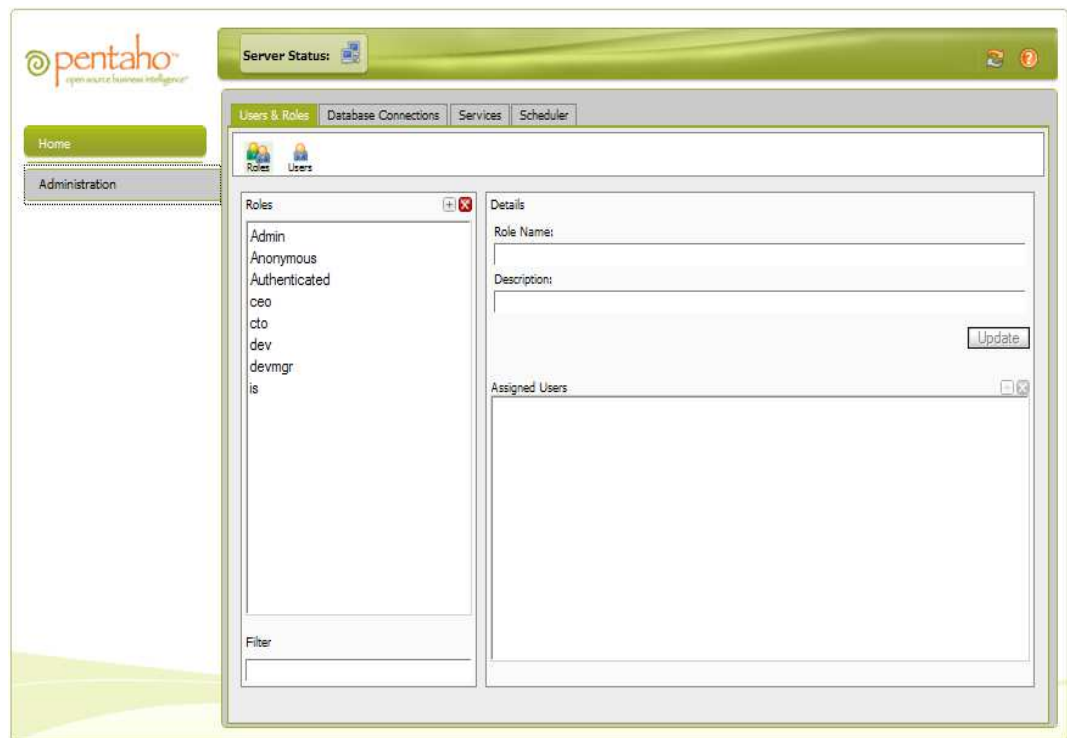


Figura 108. Pestaña de gestión de usuarios en la consola de administración

## 5.6 BI Server

EL Pentaho BI Server es el componente principal que provee Pentaho Community a los usuarios y es una aplicación Web J2EE que corre sobre un servidor de apache embebido dentro del paquete y que incluye los sub-paquetes de Pentaho Reporting,

Pentaho Analysis y Pentaho Dashboard Framework. En esta aplicación los datos son publicados, almacenados, compartidos y administrados por todos los usuarios.

Al ingresar a la aplicación web se diferencian 4 sectores en los cuales la información va a ser desplegada. La figura muestra cada uno de los sectores que se explican a continuación:

El sector uno es el escritorio de trabajo y es en donde se visualizarán los objetos creados por la herramienta (reportes, indicadores, dashboards, etc).

El sector dos es el explorador de carpetas de la plataforma y que se muestra como una estructura de árbol.

El sector tres es una vista de los archivos que se encuentran en las carpetas del segundo sector.

El sector cuatro es el menú de la aplicación y desde el mismo se puede realizar la administración básica de los componentes, así como también acciones de limpiezas de caché, cambio de idioma de la solución, acceder a la documentación, etc.



Figura 109. Sectores de trabajo de la Consola de Usuario Pentaho



## CAPITULO VI

### 6 CONCLUSIONES

#### 6.1 Resultados

Del estudio realizado para determinar la mejor herramienta en soluciones de Data Warehouse (DW) en Software Libre, tanto en base de datos como en herramientas de reportes que sirven como base para la elección de herramientas en futuros proyectos de similares características. Por otro lado se instalaron las herramientas recomendadas, base de datos, software de Extracción Transformación y Carga (ETL) y software de Inteligencia de Negocios (BI) que en este caso como resultado del análisis es Pentaho. Se desarrollaron los programas de ETL necesarios para la carga del Data Mart y se diseñaron los cubos que permiten obtener los reportes e indicadores definidos. Se implementó los Data Mart solicitados y se programó la actualización periódica del mismo de forma automatizada. Se implementaron los reportes e indicadores solicitados con la herramienta de reportes seleccionada. Se destaca como resultado obtenido la importante experiencia adquirida en temas de inteligencia de negocios a lo largo del desarrollo del proyecto.

#### 6.2 Conclusiones

Del trabajo realizado para la creación de este proyecto de tesis se pudo concluir lo siguiente:

- En la actualidad existen varias y mejores herramientas tanto libres como pagadas que se pueden utilizar en soluciones de Data Mart.
- Los datos se filtran cuando pasan desde el ambiente operacional al de depósito Data Mart. Existe mucha información que no sale del ambiente operacional, sólo los datos que realmente se necesitan ingresarán al ambiente del Data Mart.

- El desarrollo de los procesos de extracción, transformación y carga son los apropiados según la información requerida por los directores de carrera.
- Del análisis de la información se diseñaron 6 modelos dimensionales adecuados para la construcción de los indicadores académicos.
- La creación de los indicadores ayudara a tener la información al instante sin tener que pasar por un proceso largo como lo hacían hoy en día de forma manual, optimizando tiempo de respuesta y desarrollo.
- El uso de los indicadores en la facultad permitirá realizar el proceso de elaboración de las estrategias; a parte de introducirla a objetivos y medidas concretas.
- Los indicadores permiten el monitoreo del avance del cumplimiento de los objetivos estratégicos definidos y traducidos por medio de la visión de la facultad.
- El uso de una interfaz BI de reportes de Pentaho permite un manejo intuitivo y sencillo a los usuarios finales para generar sus reportes y análisis acorde a las necesidades del negocio en comparación del uso de hojas de cálculo.

### **6.3 Recomendaciones**

- El presente proyecto de tesis puede servir de base para la creación de otros sistemas que utilicen la inteligencia de negocios.
- Dedicar el tiempo necesario para el análisis de las fuentes de datos de los Data Mart, de esta manera se agilizará el trabajo al momento de la construcción y ejecución del proceso de extracción, transformación y carga.
- Tener en cuenta la compatibilidad de la herramienta que se utilizará con la base de datos ya que esta integración es la más importante para empezar a realizar el trabajo y determinar si la herramienta escogida para la solución es la más apropiada para el proyecto.

- Se recomienda recolectar la información de los requerimientos directamente con los usuarios y propietarios de la data, ya que si es a través de intermediarios dicha información puede resultar no se segura y provocara problemas en el desarrollo.
- El utilizar software libre como una herramienta para empresas medianas y pequeñas es recomendable debido a que estas no tiene los recursos para invertirlos en software propietarios ya que implicaría una inversión muy grande.
- El presente proyecto de tesis puede tener un alcance mas amplio en cuanto a la data si se contara con el apoyo para extraer la data de la fuente de información del sistema SAU.
- Generar otros reportes e indicadores académicos tanto de estudiante como docentes, utilizando los modelos dimensionales creados en este proyecto de tesis.

## **MATERIALES DE REFERENCIA**

### **BIBLIOGRAFÍA**

#### **LIBROS:**

- El Proceso de Investigación Científica, Mario Tamayo y Tamayo, 2007.
- MENDEZ, A. Carlos E.: METODOLOGIA, Guía para elaborar diseños de investigación McGraw-Hill Interamericana, S.A., Segunda edición.
- Mastering Data Warehouse Design Claudia Imhoff, Claudia Imhoff, Nicholas Galemno, Jonathan G. Geiger, 2003

#### **PÁGINAS WEB:**

- [http://www.directoriow.com/pe\\_como\\_programar\\_en\\_java\\_deitel\\_basico\\_para\\_aprender\\_java\\_17027.html](http://www.directoriow.com/pe_como_programar_en_java_deitel_basico_para_aprender_java_17027.html)<http://www.tcman.com/>
- [www.netbeans.org](http://www.netbeans.org)
- <http://planetalibro.net/ebooks/eam/index.php?label=jsp>
- <http://www.pentaho.com/>
- <http://www.businessintelligence.info/definiciones/que-es-business-intelligence.html>
- <http://dspace.esPOCH.edu.ec/handle/123456789/1519>
- <http://eduardoarea.blogspot.com/2011/10/que-son-los-cubos-olap.html>
- <http://demetole.blogspot.com/2011/10/primera-implementacion-de-pentaho-bi.html>
- <http://wiki.pentaho.com/pages/viewpage.action?pageid=14844841>
- <http://www.walkingtree.in/forums/showthread.php?548-commands-for-executing-kettle-jobs-through-command-prompt&highlight=commands>

- [http://infocenter.pentaho.com/help/index.jsp?topic=%2fpdi\\_user\\_guide%2ftopic\\_scheduling\\_scripting.html](http://infocenter.pentaho.com/help/index.jsp?topic=%2fpdi_user_guide%2ftopic_scheduling_scripting.html)
- <http://sourceforge.net/projects/mondrian/files/schema%20workbench/3.2.1-stable/>
- <http://wiki.pentaho.com/display/Reporting/Report+Designer>.
- [http://api.ning.com/files/RLZMtjtGd0kmyjoovQhnUaxUO5sC7AwcuaBAYNQ05HE\\_/howtopentaho3.5ycubomondrianengnlinux.pdf](http://api.ning.com/files/RLZMtjtGd0kmyjoovQhnUaxUO5sC7AwcuaBAYNQ05HE_/howtopentaho3.5ycubomondrianengnlinux.pdf).
- <http://community.pentaho.com/>
- <http://pentahohispano.blogspot.com/p/diccionario.html>
- <http://www.herramientasdeproductividad.com/glosario-de-inteligencia-de-negocios/>
- <http://www.gravitar.biz/index.php/bi/bi-terminologia-1/>
- <http://www.w2000mag.com/atrasados/1997/12sept97/Revista/DataMart.htm>

# ***ANEXOS***

## Anexo A.

**Presupuesto:** ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DE UN DATA MART ACADÉMICO USANDO TECNOLOGÍA DE BI PARA LA FACULTAD DE INGENIERÍA, CIENCIAS FÍSICAS Y MATEMÁTICA.

| Nº                    | RUBRO                                   | CANTIDAD | VALOR UNITARIO | VALOR RUBRO |
|-----------------------|-----------------------------------------|----------|----------------|-------------|
| 1                     | RECURSOS NECESARIOS                     |          |                |             |
|                       | • Computadores                          | 1        | 600            | 600         |
|                       | • Impresora                             | 1        | 70             | 70          |
| SUBTOTAL              |                                         |          |                | 670         |
| 2                     | RECURSOS HUMANOS                        |          |                |             |
|                       | • Tutor de Trabajo de Graduación        | 1        | ---            | ---         |
|                       | • Tribunal de Trabajo de Graduación     | 2        | ---            | ---         |
|                       | • Investigador (Autor trabajo de grado) | 1        | ---            | ---         |
| SUBTOTAL              |                                         |          |                | 0           |
| 3                     | RECURSOS MATERIALES                     |          |                |             |
|                       | Material de Escritorio                  |          |                |             |
|                       | • Toner impresora láser (B/N)           | 2        | 30             | 60          |
|                       | • Resmas de papel                       | 3        | 3,8            | 11,4        |
|                       | • Cartuchos a color CANON               | 4        | 21             | 84          |
|                       | • Caja de CDs                           | 1        | 20             | 20          |
|                       | • Carpeta de perfil                     | 4        | 0,5            | 2           |
|                       | • Empastados                            | 4        | 15             | 60          |
|                       | • Portaminas                            | 2        | 3              | 6           |
|                       | • Minas de Lápiz                        | 12       | 0,4            | 4,8         |
|                       | • Borrador                              | 2        | 0,2            | 0,4         |
|                       | Material Bibliográfico                  |          |                |             |
|                       | • Internet                              | 1 año    | 22             | 264         |
|                       | • Fotocopias de libros                  | 150      | 0,02           | 3           |
| SUBTOTAL              |                                         |          |                | 515,6       |
| 4                     | OTROS                                   |          |                |             |
|                       | • Transporte                            | 1 año    | 1              | 288         |
|                       | • Almuerzo                              | 1 año    | 1              | 660         |
|                       | • Gastos Varios                         | 1 año    | ---            | 100         |
| SUBTOTAL              |                                         |          |                | 1048        |
| TOTAL GASTOS          |                                         |          |                | 2233,6      |
| TOTAL DEL PRESUPUESTO |                                         |          |                | 2233,6      |

## Anexo B.

### CRONOGRAMA

| Id | Nombre de tarea                                | Duración | Comienzo     | Fin          |
|----|------------------------------------------------|----------|--------------|--------------|
| 1  | Fase de Planificación                          | 38 días  | lun 02/01/11 | mié 22/02/11 |
| 2  | Verificar requisitos                           | 5 días   | lun 02/01/11 | vie 06/01/11 |
| 3  | Principales requerimientos                     | 3 días   | lun 09/01/11 | mié 11/01/11 |
| 4  | Refinamiento del Plan de Desarrollo del pr     | 5 días   | jue 12/01/11 | mié 18/01/11 |
| 5  | Revisión de Presupuesto y Calendario           | 2 días   | jue 19/01/11 | vie 20/01/11 |
| 6  | Visión y Plan de desarrollo                    | 1 día    | lun 23/01/11 | lun 23/01/11 |
| 7  | Entrega de Plan de Propuesta de Tesis          | 3 días   | mar 24/01/11 | jue 26/01/11 |
| 8  | Revisión de la Plan de Tesis Y Aceptación      | 13 días  | lun 06/02/11 | mié 22/02/11 |
| 9  | Fase de Análisis de Requerimientos             | 42 días  | jue 23/02/11 | vie 20/04/11 |
| 10 | Captura de Requerimientos                      | 4 días   | jue 23/02/11 | mar 28/02/11 |
| 11 | Análisis de Requerimientos                     | 6 días   | mié 29/02/11 | mié 07/03/11 |
| 12 | Especificación de Requerimientos               | 14 días  | jue 08/03/11 | mar 27/03/11 |
| 13 | Validación de Requerimientos                   | 7 días   | mié 28/03/11 | jue 05/04/11 |
| 14 | Refinamiento de Plan Inicial                   | 5 días   | vie 06/04/11 | jue 12/04/11 |
| 15 | Revisión de la fase de Análisis de Requeri     | 6 días   | vie 13/04/11 | vie 20/04/11 |
| 16 | Fase de Elaboración                            | 66 días  | sáb 21/04/11 | lun 23/07/11 |
| 17 | Análisis de Requisitos                         | 7 días   | sáb 21/04/11 | mar 01/05/11 |
| 18 | Desarrollo de Prototipos y Casos de Usos       | 10 días  | mié 02/05/11 | mar 15/05/11 |
| 19 | Verificación de partes relevantes y críticas c | 8 días   | mié 16/05/11 | vie 25/05/11 |
| 20 | Modelo de Análisis y Diseño                    | 9 días   | sáb 26/05/11 | jue 07/06/11 |
| 21 | Definir Interfaces de Usuarios                 | 2 días   | vie 08/06/11 | lun 11/06/11 |
| 22 | Definir Arquitectura del Sistema               | 7 días   | mar 12/06/11 | mié 20/06/11 |
| 23 | Definir Esquemas de Bases de Datos             | 5 días   | jue 21/06/11 | mié 27/06/11 |
| 24 | Revisión y Aceptación del prototipo            | 6 días   | jue 28/06/11 | jue 05/07/11 |
| 25 | Revisión General                               | 2 días   | mié 06/07/11 | jue 07/07/11 |
| 26 | Asegurar el Cumplimiento de los Objetivos      | 5 días   | vie 08/07/11 | jue 14/07/11 |
| 27 | Revisión y Aceptación del Proyecto             | 6 días   | lun 16/07/11 | lun 23/07/11 |
| 28 | Fase de Construcción                           | 69 días  | mar 24/07/11 | vie 26/10/11 |
| 29 | Programación de Interfaces y Prototipos        | 11 días  | mar 24/07/11 | mar 07/08/11 |
| 30 | Elaboración y construcción del Programa        | 50 días  | mié 08/08/11 | mar 16/10/11 |
| 31 | Elaboración del Material de Apoyo              | 7 días   | mié 17/10/11 | jue 25/10/11 |
| 32 | Revisión de la Fase de Construcción            | 1 día    | vie 26/10/11 | vie 26/10/11 |
| 33 | Fase de Pruebas                                | 31 días  | lun 29/10/11 | lun 10/12/11 |
| 34 | Pruebas de Unidad                              | 11 días  | lun 29/10/11 | lun 12/11/11 |
| 35 | Pruebas de Integración                         | 7 días   | mar 13/11/11 | mié 21/11/11 |
| 36 | Pruebas de Aceptación                          | 11 días  | jue 22/11/11 | jue 06/12/11 |
| 37 | Revisión General                               | 1 día    | vie 07/12/11 | vie 07/12/11 |
| 38 | Fase de Transición                             | 35 días  | lun 10/12/11 | vie 25/01/12 |
| 39 | Entrega de documentación                       | 24 días  | lun 10/12/11 | jue 10/01/12 |
| 40 | Entrega final del producto                     | 10 días  | vie 11/01/12 | jue 24/01/12 |



## **Anexo C.**

### ***TERMINOLOGÍA BÁSICA***

#### **BI – Inteligencia de Negocios (Business Intelligence)**

Tecnología y aplicaciones de software utilizadas para reunir, poner a disposición, y manipular la información corporativa por la información efectiva, la planificación y la toma de decisiones.

#### **Data Mart**

Almacén de datos con un objetivo muy concreto normalmente limitado a un área (por ejemplo, Márketing) que se define para responder a las necesidades de un colectivo de usuarios. En otras palabras, se trata de un datawarehouse departamental, adecuado mediante transformaciones específicas para el área a la que va dirigido.

#### **Dimensión**

Perspectiva que contextualiza una medida. Suele ser un valor alfanumérico. Consiste en la agrupación de elementos con características comunes, tales como región, cliente, fecha, producto, proveedor, línea, etc.

#### **Estrategia**

La estrategia de una empresa consiste en los movimientos competitivos y enfoques de negocio que el grupo directivo de una organización emplea para hacer crecer el negocio, atraer y atender a sus clientes, competir exitosamente, conducir sus operaciones y alcanzar los niveles visualizados de desempeño organizacional.

#### **ETL (Abreviación de Extracción Transformación y Carga)**

Proceso por el que se cargan datos existentes en un sistema a otro (normalmente un Datawarehouse)

### **Granularidad**

La granularidad consiste en el nivel de detalle de la información al que decidimos descender para el análisis de los modelos. Por su parte, la multidimensionalidad nos permite analizar la información utilizando distintas dimensiones a la vez.

### **Hecho**

La tabla de hechos contiene las claves subrogadas de aquellas dimensiones que definen su nivel de detalle, y los indicadores. Nada más, por lo tanto, antes de crear la tabla de hechos debe entenderse perfectamente la información que se almacenara.

### **Medida**

Valor, generalmente numérico, que cuantifica la intersección de dimensiones. Por ejemplo, las ventas para el año 2010, en la región norte, para el cliente XYZ S.A, del producto de lujo fueron 980 mil dólares. En este caso, los 980 mil dólares corresponden a la medida de la intersección de todos los valores dimensionales (como tiempo, región, cliente y producto)

### **Modelo Entidad Relación**

Constituye una forma de representar conceptualmente la realidad basada en la representación de esta mediante su abstracción en entidades y relaciones.

De esta forma intentamos representar el mundo que nos rodea, los datos de nuestro problema mediante una serie de entidades que representan objetos o conceptos así como las relaciones que se dan entre ellos tales como su uso, composición, etc.

### **Modelo**

La representación de una porción de la realidad en sus elementos más pertinentes a la solución del problema o situación que afrontamos.

## **Multidimensionalidad**

Capacidad que ofrece una herramienta de Inteligencia de Negocios para analizar la información utilizando distintas dimensiones a la vez. De esta manera, podremos analizar, por ejemplo, cuánto hemos vendido en una determinada zona geográfica, en un mes específico, usando un canal determinado y con un margen superior al 15%.

## **Normalización**

La normalización es el proceso de organizar los datos en una base de datos. Esto incluye la creación de tablas y que establece relaciones entre aquellas tablas según reglas diseñadas para proteger los datos y hacer la base de datos que es más flexible al eliminar dos factores redundancia y Dependencia incoherente.

## **OLTP (On-Line Transactional Processing)**

Es un tipo de proceso especialmente rápido en el que las solicitudes de los usuarios son resueltas de inmediato; naturalmente, ello implica la concurrencia de un «mecanismo» que permite el procesamiento de varias transacciones a la vez.

## **OLAP (On Line Analytical Processing)**

Los sistemas de soporte a la decisión usando tecnologías de Data Warehouse, se llaman sistemas OLAP. En general, estos sistemas deben: Soportar requerimientos complejos de análisis, Analizar datos desde diferentes perspectivas, Soportar análisis complejos contra un volumen ingente de datos

La funcionalidad de los sistemas OLAP se caracteriza por ser un análisis multidimensional de datos corporativos, que soportan los análisis del usuario y unas posibilidades de navegación, seleccionando la información a obtener.

## **OLAP vs OLTP**

OLAP o procesamiento analítico en línea, es el paradigma de análisis multidimensional de un Data Warehouse. Este concepto se contrapone al de OLTP

o procesamiento transaccional en línea, que es el empleado por los sistemas fuentes para optimizar la recogida de información. En pocas palabras, mientras OLTP se encarga de procesar óptimamente multitud de pequeñas transacciones de captura de información (su entrada, modificación o eliminación), OLAP se dedica al análisis de enormes cantidades de información; por ende realizará pocas transacciones, pero éstas procesaran un volumen muy superior (cientos de miles de registros).

### **MOLAP**

La arquitectura MOLAP usa unas bases de datos multidimensionales para proporcionar el análisis, su principal premisa es que el OLAP está mejor implantado almacenando los datos multidimensionalmente.

### **ROLAP**

La arquitectura ROLAP cree que las capacidades OLAP están perfectamente implantadas sobre bases de datos relacionales la arquitectura ROLAP es capaz de usar datos precalculados si estos están disponibles, o de generar dinámicamente los resultados desde los datos elementales si es preciso. Esta arquitectura accede directamente a los datos del Data Warehouse, y soporta técnicas de optimización de accesos para acelerar las consultas.

### **MDX** (Multi Dimensional Expressions)

Es el lenguaje de consulta para recuperar datos de cubos OLAP No existe ningún estándar todavía. Y ya que lo inventó Microsoft pues se toma como base el trabajo que desarrollaron ellos.

### **Mondrian**

Motor OLAP Open Source. Forma parte de la suite Pentaho aunque se puede encontrar en multitud de otros proyectos. Es junto con Palo el gran motor OLAP open source.

**Anexo D.****MAPEO DE DIMENSIONES****DIM\_FECHA**

| DESCRIPCION                                        | DWH               | TIPO DATO |
|----------------------------------------------------|-------------------|-----------|
| Clave subrogada de la fecha                        | SK_FECHA          | Number    |
| Fecha en formato 1999-01-01                        | FEC_ID_FECHA      | Date      |
| Fecha en formato 1 de enero de 1999                | FEC_DESC_FECHA    | Varchar   |
| Número del día de la semana                        | FEC_ID_DIASEM     | Number    |
| Descripción del día de la semana                   | FEC_DESC_DIASEM   | Varchar   |
| Número del día del mes                             | FEC_ID_DIAMES     | Number    |
| Número del día del año                             | FEC_ID_DIAANIO    | Number    |
| Año concatenado con el número del día de la semana | FEC_ID_SEMANA     | Varchar   |
| Semana concatenado el número de la semana y el año | FEC_DESC_SEMANA   | Varchar   |
| Número de la semana del año                        | FEC_ID_SEMANAN    | Number    |
| Semana concatenado el número de la semana          | FEC_DESC_SEMANAN  | Varchar   |
| Si es festivo o no                                 | FEC_ID_ES_FESTIVO | Varchar   |
| Si es fin de semana o no                           | FEC_ID_ES_FINDE   | Varchar   |
| Año concatenado el mes                             | FEC_ID_MES        | Varchar   |
| Mes concatenado el año                             | FEC_DESC_MES      | Varchar   |
| Número del mes                                     | FEC_ID_MESN       | Number    |
| Descripción del mes                                | FEC_DESC_MESN     | Varchar   |
| Año concatenado el trimestre                       | FEC_ID_TRIM       | Number    |
| Trimestre concatenado con el año                   | FEC_DESC_TRIM     | Varchar   |
| Número del trimestre                               | FEC_ID_TRIMN      | Number    |
| Número del trimestre concatenado a Trim            | FEC_DESC_TRIMN    | Varchar   |
| Número del semestre                                | FEC_ID_SEMESTRE   | Number    |
| Número del semestre concatenado el año             | FEC_DESC_SEMESTRE | Varchar   |
| Año                                                | FEC_ID_ANIO       | Number    |
| Año anterior                                       | FEC_ID_ANIO_ANT   | Number    |
| Fecha ultimo día del mes                           | FEC_ULTIMO_DIA    | Varchar   |
| Código si es ultimo día o no                       | FEC_ULTIMON_DIA   | Number    |

**Tabla 3. Mapeo DIM\_FECHA**

## DIM\_PERIODO\_LECTIVO

| DESCRIPCION                 | DWH            | TIPO DATO | SYBASE     | STG_DWH    |
|-----------------------------|----------------|-----------|------------|------------|
| Clave subrogada del periodo | SK_PERIODO     | Number    |            |            |
| Código del periodo          | PL_COD_PERIODO | Number    | PER_CODIGO | PER_CODIGO |
| Descripción del periodo     | PL_DES_PERIODO | Varchar   | PER_NOMBRE | PER_NOMBRE |

Tabla 4. Mapeo DIM\_PERIODO\_LECTIVO

## DIM\_CATEGORIA\_DOCENTE

| DESCRIPCION                     | DWH              | TIPO DATO | SYBASE            | STG_DWH           |
|---------------------------------|------------------|-----------|-------------------|-------------------|
| Clave subrogada de la categoría | SK_CATEGORIA     | Number    |                   |                   |
| Código de la categoría          | CD_COD_CATEGORIA | Varchar   | CON_CATEGORIA     | CON_CATEGORIA     |
| Descripción de la categoría     | CD_DES_CATEGORIA | Varchar   | CON_DES_CATEGORIA | CON_DES_CATEGORIA |

Tabla 5. Mapeo DIM\_CATEGORIA\_DOCENTE

## DIM\_DEDICACION\_DOCENTE

| DESCRIPCION                      | DWH               | TIPO DATO | SYBASE              | STG_DWH             |
|----------------------------------|-------------------|-----------|---------------------|---------------------|
| Clave subrogada de la dedicación | SK_DEDICACION     | Number    |                     |                     |
| Código de dedicación             | DD_COD_DEDICACION | Varchar   | CON_DEDICACION      | CON_DEDICACION      |
| Descripción de dedicación        | DD_DES_DEDICACION | Varchar   | CON_DESC_DEDICACION | CON_DESC_DEDICACION |
| Carga horaria                    | DD_CARGA_HORARIA  | Number    | CON_CARGA_HORARIA   | CON_CARGA_HORARIA   |

Tabla 6. Mapeo DIM\_DEDICACION\_DOCENTE

## DIM\_ESTADO\_ESTUDIANTE

| DESCRIPCION                               | DWH           | TIPO DATO | SYBASE     | STG_DWH    |
|-------------------------------------------|---------------|-----------|------------|------------|
| Clave subrogada del estado del estudiante | SK_ESTADO     | Number    |            |            |
| Código del Estado                         | EE_COD_ESTADO | Number    | EST_ESTADO | EST_ESTADO |
| Descripción del Estado                    | EE_DES_ESTADO | Varchar   |            | EST_NOMBRE |

Tabla 7. Mapeo DIM\_ESTADO\_ESTUDIANTE

## DIM\_FACULTAD

| DESCRIPCION                    | DWH             | TIPO DATO | SYBASE     | STG_DWH    |
|--------------------------------|-----------------|-----------|------------|------------|
| Clave subrogada de la facultad | SK_FACULTAD     | Number    |            |            |
| Código de la facultad          | FA_COD_FACULTAD | Number    | FAC_CODIGO | FAC_CODIGO |
| Descripción de la facultad     | FA_DES_FACULTAD | Varchar   | FAC_NOMBRE | FAC_NOMBRE |
| Código de la escuela           | FA_COD_ESCUELA  | Number    | ESC_CODIGO | ESC_CODIGO |
| Descripción de la escuela      | FA_DES_ESCUELA  | Varchar   | ESC_NOMBRE | ESC_NOMBRE |
| Código de Carrera              | FA_COD_CARRERA  | Number    | ESP_CODIGO | ESP_CODIGO |
| Descripción de Carrera         | FA_DES_CARRERA  | Varchar   | ESP_NOMBRE | ESP_NOMBRE |
| Código de la Materia           | FA_COD_MATERIA  | Number    | MAT_CODIGO | MAT_CODIGO |
| Descripción de la Materia      | FA_DES_MATERIA  | Varchar   | MAT_NOMBRE | MAT_NOMBRE |

Tabla 8. Mapeo DIM\_FACULTAD

## DIM\_ESPECIALIDAD

| DESCRIPCION                    | DWH             | TIPO DATO | SYBASE     | STG_DWH    |
|--------------------------------|-----------------|-----------|------------|------------|
| Clave subrogada de la facultad | SK_ESPECIALIDAD | Number    |            |            |
| Código de la facultad          | SP_COD_FACULTAD | Number    | FAC_CODIGO | FAC_CODIGO |
| Descripción de la facultad     | SP_DES_FACULTAD | Varchar   | FAC_NOMBRE | FAC_NOMBRE |
| Código de la escuela           | SP_COD_ESCUELA  | Number    | ESC_CODIGO | ESC_CODIGO |
| Descripción de la escuela      | SP_DES_ESCUELA  | Varchar   | ESC_NOMBRE | ESC_NOMBRE |
| Código de Carrera              | SP_COD_CARRERA  | Number    | ESP_CODIGO | ESP_CODIGO |
| Descripción de Carrera         | SP_DES_CARRERA  | Varchar   | ESP_NOMBRE | ESP_NOMBRE |

Tabla 9. Mapeo DIM\_ESPECIALIDAD

### **DIM\_ESTADO\_ESTU\_CURSO**

| DESCRIPCION                                 | DWH             | TIPO DATO | SYBASE     | STG_DWH         |
|---------------------------------------------|-----------------|-----------|------------|-----------------|
| Clave subrogada del estado estudiante curso | SK_ESTADO_CURSO | Number    |            |                 |
| Código estado curso                         | EC_COD_ESTCURSO | Varchar   | INS_ESTADO | INS_ESTADO      |
| Descripción estado curso                    | EC_DES_ESTCURSO | Varchar   |            | INS_DESC_ESTADO |

Tabla 10. Mapeo DIM\_ESTADO\_ESTU\_CURSO

### **DIM\_CURSO**

| DESCRIPCION               | DWH          | TIPO DATO | SYBASE     | STG_DWH         |
|---------------------------|--------------|-----------|------------|-----------------|
| Clave subrogada del curso | SK_CURSO     | Number    |            |                 |
| Código curso              | CR_COD_CURSO | Varchar   | INS_ESTADO | INS_ESTADO      |
| Descripción curso         | CR_DES_CURSO | Varchar   |            | INS_DESC_ESTADO |

Tabla 11. Mapeo DIM\_CURSO

### **DIM\_PROFESOR**

| DESCRIPCION                  | DWH             | TIPO DATO | SYBASE          | STG_DWH         |
|------------------------------|-----------------|-----------|-----------------|-----------------|
| Clave subrogada del profesor | SK_PROFESOR     | Number    |                 |                 |
| Código del Profesor          | PR_COD_PROFESOR | Number    | PROF_CODIGO     | PROF_CODIGO     |
| Descripción del Profesor     | PR_DES_PROFESOR | Varchar   | NOMBRE_PROFESOR | NOMBRE_PROFESOR |
| Cédula del Profesor          | PR_CEDULA       | Varchar   | PROF_CEDULA_ID  | PROF_CEDULA_ID  |

Tabla 12. Mapeo DIM\_PROFESOR



## DIM\_ESTUDIANTE

| DESCRIPCIÓN                    | DWH                    | TIPO DATO | SYBASE                     | STG_DWH                    |
|--------------------------------|------------------------|-----------|----------------------------|----------------------------|
| Clave subrogada del estudiante | SK_ESTUDIANTE          | Number    |                            |                            |
| Código del Estudiante          | EST_COD_ESTUDIANTE     | Number    | EST_CODIGO                 | EST_CODIGO                 |
| Descripción del Estudiante     | EST_DES_ESTUDIANTE     | Varchar   | EST_NOMBRE<br>EST_APELLIDO | EST_NOMBREC                |
| Cédula del Estudiante          | EST_CEDULA             | Varchar   | EST_CEDULA                 | EST_CEDULA                 |
| Código Sexo                    | EST_SEXO               | Varchar   | EST_SEXO                   | EST_SEXO                   |
| Descripción Sexo               | EST_SEXO_DESC          | Varchar   | EST_SEXO                   | EST_SEXO_DESC              |
| Código Modalidad               | EST_MODALIDAD          | Number    | EST_MODALIDAD              | EST_MODALIDAD              |
| Descripción Modalidad          | EST_MODALIDAD_DESC     | Varchar   |                            | Valor quemado 'Presencial' |
| Fecha de Nacimiento            | EST_FECHA_NACIMIENTO   | Date      | EST_FECHA_NACIMIENTO       | EST_FECHA_NACIMIENTO       |
| Nota grado colegio             | EST_NOTA_GRADO_COLEGIO | Number    | EST_NOTA_GRADO_COLEGIO     | EST_NOTA_GRADO_COLEGIO     |
| Código País                    | EST_CODIGO_PAIS        | Number    | PAI_CODIGO                 | PAI_CODIGO                 |
| Descripción País               | EST_DES_PAIS           | Varchar   | PAI_NOMBRE                 | PAI_NOMBRE                 |
| Código Ciudad                  | EST_CODIGO_CIUDAD      | Number    | CIU_CODIGO                 | CIU_CODIGO                 |
| Descripción Ciudad             | EST_DES_CIUDAD         | Varchar   | CIU_NOMBRE                 | CIU_NOMBRE                 |
| Código Colegio                 | EST_CODIGO_COLEGIO     | Number    | COL_CODIGO                 | COL_NOMBRE                 |
| Descripción Colegio            | EST_DES_COLEGIO        | Varchar   | COL_NOMBRE                 | COL_NOMBRE                 |
| Código Especialidad            | EST_CODIGO_ESP         | Number    | COL_ESP_CODIGO             | COL_ESP_CODIGO             |
| Descripción Especialidad       | EST_DES_ESP            | Varchar   | ESP_NOMBRE                 | ESP_NOMBRE                 |
| Código Nacionalidad            | EST_CODIGO_NAC         | Number    | NAC_CODIGO                 | NAC_CODIGO                 |
| Descripción Nacionalidad       | EST_DES_NAC            | Varchar   | NAC_NOMBRE                 | NAC_NOMBRE                 |

Tabla 13. Mapeo DIM\_ESTUDIANTE

## Anexo E.

### MAPEO DE HECHOS

#### FC\_DOCENTE\_DET

| DESCRIPCION                                      | CAMPO             | TIPO DATO | SYBASE                                | STG_DWH           |
|--------------------------------------------------|-------------------|-----------|---------------------------------------|-------------------|
| Clave subrogada en base al código del periodo    | SK_PERIODO        | Number    |                                       |                   |
| Clave subrogada en base al código del categoría  | SK_CATEGORIA      | Number    |                                       |                   |
| Clave subrogada en base al dedicación            | SK_DEDICACION     | Number    |                                       |                   |
| Clave subrogada en base al código de la facultad | SK_FACULTAD       | Number    |                                       |                   |
| Clave subrogada en base al código del profesor   | SK_PROFESOR       | Number    |                                       |                   |
| Clave subrogada en base al código del curso      | SK_CURSO          | Number    |                                       |                   |
| Código del periodo                               | DC_COD_PERIODO    | Number    | PER_CODIGO                            | PER_CODIGO        |
| Código del profesor                              | DC_COD_PROFESOR   | Number    | PROF_CODIGO                           | PROF_CODIGO       |
| Código de la dedicación                          | DC_COD_DEDICACION | Varchar   | CON_DEDICACION                        | CON_DEDICACION    |
| Código de la categoría                           | DC_COD_CATEGORIA  | Varchar   | CON_CATEGORIA                         | CON_CATEGORIA     |
| Código de la carrera                             | DC_COD_CARRERA    | Number    | ESC_CODIGO                            | ESC_CODIGO        |
| Hora inicio                                      | DC_HORA_INICIO    | Date      | HORA_HORA_INICIO                      | HORA_HORA_INICIO  |
| Hora final                                       | DC_HORA_FINAL     | Date      | HORA_HORA_FINAL                       | HORA_HORA_FINAL   |
| Hora clase                                       | DC_HORA_CLASE     | Number    | HORA_HORA_FINAL -<br>HORA_HORA_INICIO |                   |
| Código día                                       | DC_COD_DIA        | Number    | HORA_CODIGO_DIA                       | HORA_CODIGO_DIA   |
| Descripción del día                              | DC_DESC_DIA       | Varchar   | Calculo                               |                   |
| Descripción del aula                             | DC_AULA_NOMBRE    | Varchar   | AULA_NOMBRE                           | AULA_NOMBRE       |
| Código del curso                                 | DC_COD_CURSO      | Number    | CUR_CODIGO                            | CUR_CODIGO        |
| Descripción del paralelo                         | DC_MAT_PARALELO   | Number    | MATXPROF_PARALELO                     | MATXPROF_PARALELO |

|                      |                |        |                 |            |
|----------------------|----------------|--------|-----------------|------------|
| Código de la materia | DC_COD_MATERIA | Number | MAT_NOMBRE      | MAT_NOMBRE |
| Contador             | DC_CONTADOR    | Number | Valor quemado 1 |            |

Tabla 14. Mapeo FC\_DOCENTE\_DET

### FC\_DOCENTE

| DESCRIPCION                                      | CAMPO          | TIPO DATO | SYBASE                                | STG_DWH          |
|--------------------------------------------------|----------------|-----------|---------------------------------------|------------------|
| Clave subrogada en base al código del periodo    | SK_PERIODO     | Number    |                                       |                  |
| Clave subrogada en base al código del categoría  | SK_CATEGORIA   | Number    |                                       |                  |
| Clave subrogada en base al dedicación            | SK_DEDICACION  | Number    |                                       |                  |
| Clave subrogada en base al código de la facultad | SK_FACULTAD    | Number    |                                       |                  |
| Clave subrogada en base al código del profesor   | SK_PROFESOR    | Number    |                                       |                  |
| Clave subrogada en base al código del curso      | SK_CURSO       | Number    |                                       |                  |
| Hora inicio                                      | DC_HORA_INICIO | Date      | HORA_HORA_INICIO                      | HORA_HORA_INICIO |
| Hora final                                       | DC_HORA_FINAL  | Date      | HORA_HORA_FINAL                       | HORA_HORA_FINAL  |
| Hora clase                                       | DC_HORA_CLASE  | Number    | HORA_HORA_FINAL -<br>HORA_HORA_INICIO |                  |
| Código día                                       | DC_COD_DIA     | Number    | HORA_CODIGO_DIA                       | HORA_CODIGO_DIA  |

|                          |                 |         |                   |                   |
|--------------------------|-----------------|---------|-------------------|-------------------|
| Descripción del día      | DC_DESC_DIA     | Varchar | Cálculo           |                   |
| Descripción del aula     | DC_AULA_NOMBRE  | Varchar | AULA_NOMBRE       | AULA_NOMBRE       |
| Descripción del paralelo | DC_MAT_PARALELO | Number  | MATXPROF_PARALELO | MATXPROF_PARALELO |
| Contador                 | DC_CONTADOR     | Number  | Valor quemado 1   |                   |
| Total                    | DC_TOTAL        | Number  | Cálculo           |                   |

**Tabla 15.** Mapeo FC\_DOCENTE

### **FC\_ESTUDIANTE\_DET**

| DESCRIPCION                                                 | CAMPO           | TIPO DATO | SYBASE               | STG_DWH              |
|-------------------------------------------------------------|-----------------|-----------|----------------------|----------------------|
| Clave subrogada en base al código del periodo               | SK_PERIODO      | Number    |                      |                      |
| Clave subrogada en base al código de la especialidad        | SK_ESPECIALIDAD | Number    |                      |                      |
| Clave subrogada en base al código del profesor              | SK_ESTUDIANTE   | Number    |                      |                      |
| Clave subrogada en base al código del curso                 | SK_CURSO        | Number    |                      |                      |
| Clave subrogada en base al código del estado del estudiante | SK_ESTADO_CURSO | Number    |                      |                      |
| Clave subrogada en base al código de la fecha de insc.      | SK_FECHA        | Number    |                      |                      |
| Fecha                                                       | ST_FECHA        | Date      | EST_FECHA_NACIMIENTO | EST_FECHA_NACIMIENTO |
| Código de la carrera                                        | ST_COD_CARRERA  | Number    | ESC_CODIGO           | ESC_CODIGO           |

|                                  |                        |         |                         |                         |
|----------------------------------|------------------------|---------|-------------------------|-------------------------|
| Código del periodo               | ST_COD_PERIODO         | Number  | PER_CODIGO              | PER_CODIGO              |
| Código del estudiante            | ST_COD_ESTUDIANTE      | Number  | EST_CODIGO              | EST_CODIGO              |
| Código del curso                 | ST_COD_CURSO           | Number  | CUR_CODIGO              | CUR_CODIGO              |
| Código del estado del estudiante | ST_COD_ESTCURSO        | Varchar | INS_ESTADO              | INS_ESTADO              |
| Número matrícula curso           | ST_NUM_MATRICULA_CURSO | Number  | INS_NUM_MATRICULA_CURSO | INS_NUM_MATRICULA_CURSO |
| Contador                         | ST_CONTADOR            | Number  | Valor quemado 1         |                         |
| Edad                             | ST_EDAD                | Number  | EST_EDAD                | EST_EDAD                |
| Estado Civil                     | ST_ESTADO_CIVIL        | Varchar | EST_ESTADO_CIVIL        | EST_ESTADO_CIVIL        |
| Descripción estado civil         | ST_DES_ESTADO_CIVIL    | Varchar | Cálculo                 |                         |
| Estado Trabajo                   | ST_ESTADO_TRABAJO      | Varchar | EST_TRABAJO_ESTADO      | EST_TRABAJO_ESTADO      |
| Descripción estado trabajo       | ST_DES_ESTADO_TRABAJO  | Varchar | Cálculo                 |                         |

Tabla 16. Mapeo FC\_ESTUDIANTE\_DET

## FC\_ESTUDIANTE

| DESCRIPCION                                          | CAMPO           | TIPO DATO | SYBASE | STG_DWH |
|------------------------------------------------------|-----------------|-----------|--------|---------|
| Clave subrogada en base al código del periodo        | SK_PERIODO      | Number    |        |         |
| Clave subrogada en base al código de la especialidad | SK_ESPECIALIDAD | Number    |        |         |
| Clave subrogada en base al código del profesor       | SK_ESTUDIANTE   | Number    |        |         |

|                                                              |                        |         |                         |                         |
|--------------------------------------------------------------|------------------------|---------|-------------------------|-------------------------|
| Clave subrogada en base al código del curso                  | SK_CURSO               | Number  |                         |                         |
| Clave subrogada en base al código del estado del estudiante  | SK_ESTADO_CURSO        | Number  |                         |                         |
| Clave subrogada en base al código de la fecha de inscripción | SK_FECHA               | Number  |                         |                         |
| Número matrícula curso                                       | ST_NUM_MATRICULA_CURSO | Number  | INS_NUM_MATRICULA_CURSO | INS_NUM_MATRICULA_CURSO |
| Contador                                                     | ST_CONTADOR            | Number  | Valor quemado 1         |                         |
| Edad                                                         | ST_EDAD                | Number  | EST_EDAD                | EST_EDAD                |
| Estado Civil                                                 | ST_ESTADO_CIVIL        | Varchar | EST_ESTADO_CIVIL        | EST_ESTADO_CIVIL        |
| Descripción estado civil                                     | ST_DES_ESTADO_CIVIL    | Varchar | Cálculo                 |                         |
| Estado Trabajo                                               | ST_ESTADO_TRABAJO      | Varchar | EST_TRABAJO_ESTADO      | EST_TRABAJO_ESTADO      |
| Descripción estado trabajo                                   | ST_DES_ESTADO_TRABAJO  | Varchar | Cálculo                 |                         |

Tabla 17. Mapeo FC\_ESTUDIANTE

### FC\_REG\_ACADEMICO\_DET

| DESCRIPCION                                   | CAMPO      | TIPO DATO | SYBASE | STG_DWH |
|-----------------------------------------------|------------|-----------|--------|---------|
| Clave subrogada en base al código del periodo | SK_PERIODO | Number    |        |         |

|                                                      |                   |        |                         |                         |
|------------------------------------------------------|-------------------|--------|-------------------------|-------------------------|
| Clave subrogada en base al código de la especialidad | SK_ESPECIALIDAD   | Number |                         |                         |
| Clave subrogada en base al código del profesor       | SK_ESTUDIANTE     | Number |                         |                         |
| Clave subrogada en base al código del curso          | SK_CURSO          | Number |                         |                         |
| Clave subrogada del estado del estudiante            | SK_ESTADO_CURSO   | Number |                         |                         |
| Clave subrogada de la fecha                          | SK_FECHA          | Number |                         |                         |
| Código del periodo                                   | RA_COD_PERIODO    | Number | PER_CODIGO              | PER_CODIGO              |
| Código de la carrera                                 | RA_COD_CARRERA    | Number | ESP_CODIGO              | ESP_CODIGO              |
| Código del estudiante                                | RA_COD_ESTUDIANTE | Number | EST_CODIGO              | EST_CODIGO              |
| Código del curso                                     | RA_COD_CURSO      | Number | CUR_CODIGO              | CUR_CODIGO              |
| Número matrícula curso                               | RA_NUM_MATRICULA  | Number | INS_NUM_MATRICULA_CURSO | INS_NUM_MATRICULA_CURSO |
| Código del estado                                    | RA_COD_ESTCURSO   | Number | INS_ESTADO              | INS_ESTADO              |
| Fecha inscripción estudiante                         | RA_FECHA          | Date   | INS_FECHA               | INS_FECHA               |
| Contador                                             | RA_CONTADOR       | Number | Valor quemado 1         |                         |

Tabla 18. Mapeo FC\_REG\_ACADEMICO\_DET

## FC\_REGISTRO\_ACADEMICO

| DESCRIPCION                                          | CAMPO            | TIPO DATO | SYBASE                  | STG_DWH                 |
|------------------------------------------------------|------------------|-----------|-------------------------|-------------------------|
| Clave subrogada en base al código del periodo        | SK_PERIODO       | Number    |                         |                         |
| Clave subrogada en base al código de la especialidad | SK_ESPECIALIDAD  | Number    |                         |                         |
| Clave subrogada en base al código del profesor       | SK_ESTUDIANTE    | Number    |                         |                         |
| Clave subrogada en base al código del curso          | SK_CURSO         | Number    |                         |                         |
| Clave subrogada del estado del estudiante            | SK_ESTADO_CURSO  | Number    |                         |                         |
| Clave subrogada de la fecha                          | SK_FECHA         | Number    |                         |                         |
| Número matrícula curso                               | RA_NUM_MATRICULA | Number    | INS_NUM_MATRICULA_CURSO | INS_NUM_MATRICULA_CURSO |
| Contador                                             | RA_CONTADOR      | Number    | Valor quemado 1         |                         |

Tabla 19. Mapeo FC\_REGISTRO\_ACADEMICO



## FC\_COHORTE\_DET

| DESCRIPCION                                          | CAMPO             | TIPO DATO | SYBASE                  | STG_DWH                 |
|------------------------------------------------------|-------------------|-----------|-------------------------|-------------------------|
| Clave subrogada de la fecha                          | SK_FECHA          | Number    |                         |                         |
| Clave subrogada en base al código del periodo        | SK_PERIODO        | Number    |                         |                         |
| Clave subrogada en base al código de la especialidad | SK_ESPECIALIDAD   | Number    |                         |                         |
| Clave subrogada en base al código del profesor       | SK_ESTUDIANTE     | Number    |                         |                         |
| Clave subrogada en base al código del curso          | SK_CURSO          | Number    |                         |                         |
| Clave subrogada del estado del estudiante            | SK_ESTADO_CURSO   | Number    |                         |                         |
| Código del periodo                                   | CH_COD_PERIODO    | Number    | PER_CODIGO              | PER_CODIGO              |
| Código de la carrera                                 | CH_COD_CARRERA    | Number    | ESP_CODIGO              | ESP_CODIGO              |
| Código del estudiante                                | CH_COD_ESTUDIANTE | Number    | EST_CODIGO              | EST_CODIGO              |
| Código del curso                                     | CH_COD_CURSO      | Number    | CUR_CODIGO              | CUR_CODIGO              |
| Número matrícula curso                               | CH_NUM_MATRICULA  | Number    | INS_NUM_MATRICULA_CURSO | INS_NUM_MATRICULA_CURSO |
| Código del estado                                    | CH_COD_ESTCURSO   | Number    | INS_ESTADO              | INS_ESTADO              |
| Fecha inscripción estudiante                         | CH_FECHA          | Date      | INS_FECHA               | INS_FECHA               |
| Contador                                             | CH_CONTADOR       | Number    | Valor quemado 1         |                         |

Tabla 20. Mapeo FC\_COHORTE\_DET

## FC\_COHORTE

| DESCRIPCION                                          | CAMPO            | TIPO DATO | SYBASE                  | STG_DWH                 |
|------------------------------------------------------|------------------|-----------|-------------------------|-------------------------|
| Clave subrogada de la fecha                          | SK_FECHA         | Number    |                         |                         |
| Clave subrogada en base al código del periodo        | SK_PERIODO       | Number    |                         |                         |
| Clave subrogada en base al código de la especialidad | SK_ESPECIALIDAD  | Number    |                         |                         |
| Clave subrogada en base al código del profesor       | SK_ESTUDIANTE    | Number    |                         |                         |
| Clave subrogada en base al código del curso          | SK_CURSO         | Number    |                         |                         |
| Clave subrogada del estado del estudiante            | SK_ESTADO_CURSO  | Number    |                         |                         |
| Número matrícula curso                               | CH_NUM_MATRICULA | Number    | INS_NUM_MATRICULA_CURSO | INS_NUM_MATRICULA_CURSO |
| Contador                                             | CH_CONTADOR      | Number    | Valor quemado 1         |                         |

Tabla 21. Mapeo FC\_COHORTE

## FC\_DESERCION\_DET

| DESCRIPCION                                          | CAMPO             | TIPO DATO | DWH            |
|------------------------------------------------------|-------------------|-----------|----------------|
| Clave subrogada en base al código del periodo        | SK_PERIODO        | Number    |                |
| Clave subrogada en base al código de la especialidad | SK_ESPECIALIDAD   | Number    |                |
| Descripción de la cohorte                            | DE_DES_COHORTE    | Varchar   | Valor quemado  |
| Código del periodo                                   | DE_COD_PERIODO    | Number    | RA_COD_PERIODO |
| Código de la carrera                                 | DE_COD_CARRERA    | Number    | RA_COD_CARRERA |
| Número de estudiantes                                | NUM_ESTUDIANTES   | Number    | CONTADOR       |
| Periodo                                              | DE_DET_PERIOSO_A  | Varchar   | Valor quemado  |
| Código periodo                                       | DE_COD_PERIODO_A  | Number    | RA_COD_PERIODO |
| Descripción periodo                                  | DE_DES_PERIODO_A  | Varchar   | RA_DES_PERIODO |
| Código carrera                                       | DE_COD_CARRERA_A  | Number    | RA_COD_CARRERA |
| Número de estudiantes                                | NUM_ESTUDIANTES_A | Number    | CONTADOR       |
| Periodo                                              | DE_DET_PERIOSO_B  | Varchar   | Valor quemado  |
| Código periodo                                       | DE_COD_PERIODO_B  | Number    | RA_COD_PERIODO |
| Descripción periodo                                  | DE_DES_PERIODO_B  | Varchar   | RA_DES_PERIODO |
| Código carrera                                       | DE_COD_CARRERA_B  | Number    | RA_COD_CARRERA |
| Número de estudiantes                                | NUM_ESTUDIANTES_B | Number    | CONTADOR       |
| Periodo                                              | DE_DET_PERIOSO_C  | Varchar   | Valor quemado  |
| Código periodo                                       | DE_COD_PERIODO_C  | Number    | RA_COD_PERIODO |
| Descripción periodo                                  | DE_DES_PERIODO_C  | Varchar   | RA_DES_PERIODO |
| Código carrera                                       | DE_COD_CARRERA_C  | Number    | RA_COD_CARRERA |
| Número de estudiantes                                | NUM_ESTUDIANTES_C | Number    | CONTADOR       |

Tabla 22. Mapeo FC\_DESERCION\_DET

## FC\_DESERCION

| DESCRIPCION                                          | CAMPO             | TIPO DATO | DWH            |
|------------------------------------------------------|-------------------|-----------|----------------|
| Clave subrogada en base al código del periodo        | SK_PERIODO        | Number    |                |
| Clave subrogada en base al código de la especialidad | SK_ESPECIALIDAD   | Number    |                |
| Descripción de la cohorte                            | DE_DES_COHORTE    | Varchar   | Valor quemado  |
| Código del periodo                                   | DE_COD_PERIODO    | Number    | RA_COD_PERIODO |
| Código de la carrera                                 | DE_COD_CARRERA    | Number    | RA_COD_CARRERA |
| Número de estudiantes                                | NUM_ESTUDIANTES   | Number    | CONTADOR       |
| Periodo                                              | DE_DET_PERIODO_A  | Varchar   | Valor quemado  |
| Código periodo                                       | DE_COD_PERIODO_A  | Number    | RA_COD_PERIODO |
| Descripción periodo                                  | DE_DES_PERIODO_A  | Varchar   | RA_DES_PERIODO |
| Código carrera                                       | DE_COD_CARRERA_A  | Number    | RA_COD_CARRERA |
| Número de estudiantes                                | NUM_ESTUDIANTES_A | Number    | CONTADOR       |
| Periodo                                              | DE_DET_PERIODO_B  | Varchar   | Valor quemado  |
| Código periodo                                       | DE_COD_PERIODO_B  | Number    | RA_COD_PERIODO |
| Descripción periodo                                  | DE_DES_PERIODO_B  | Varchar   | RA_DES_PERIODO |
| Código carrera                                       | DE_COD_CARRERA_B  | Number    | RA_COD_CARRERA |
| Número de estudiantes                                | NUM_ESTUDIANTES_B | Number    | CONTADOR       |
| Periodo                                              | DE_DET_PERIODO_C  | Varchar   | Valor quemado  |
| Código periodo                                       | DE_COD_PERIODO_C  | Number    | RA_COD_PERIODO |
| Descripción periodo                                  | DE_DES_PERIODO_C  | Varchar   | RA_DES_PERIODO |
| Código carrera                                       | DE_COD_CARRERA_C  | Number    | RA_COD_CARRERA |
| Número de estudiantes                                | NUM_ESTUDIANTES_C | Number    | CONTADOR       |

Tabla 23. Mapeo FC\_DESERCION\_DET